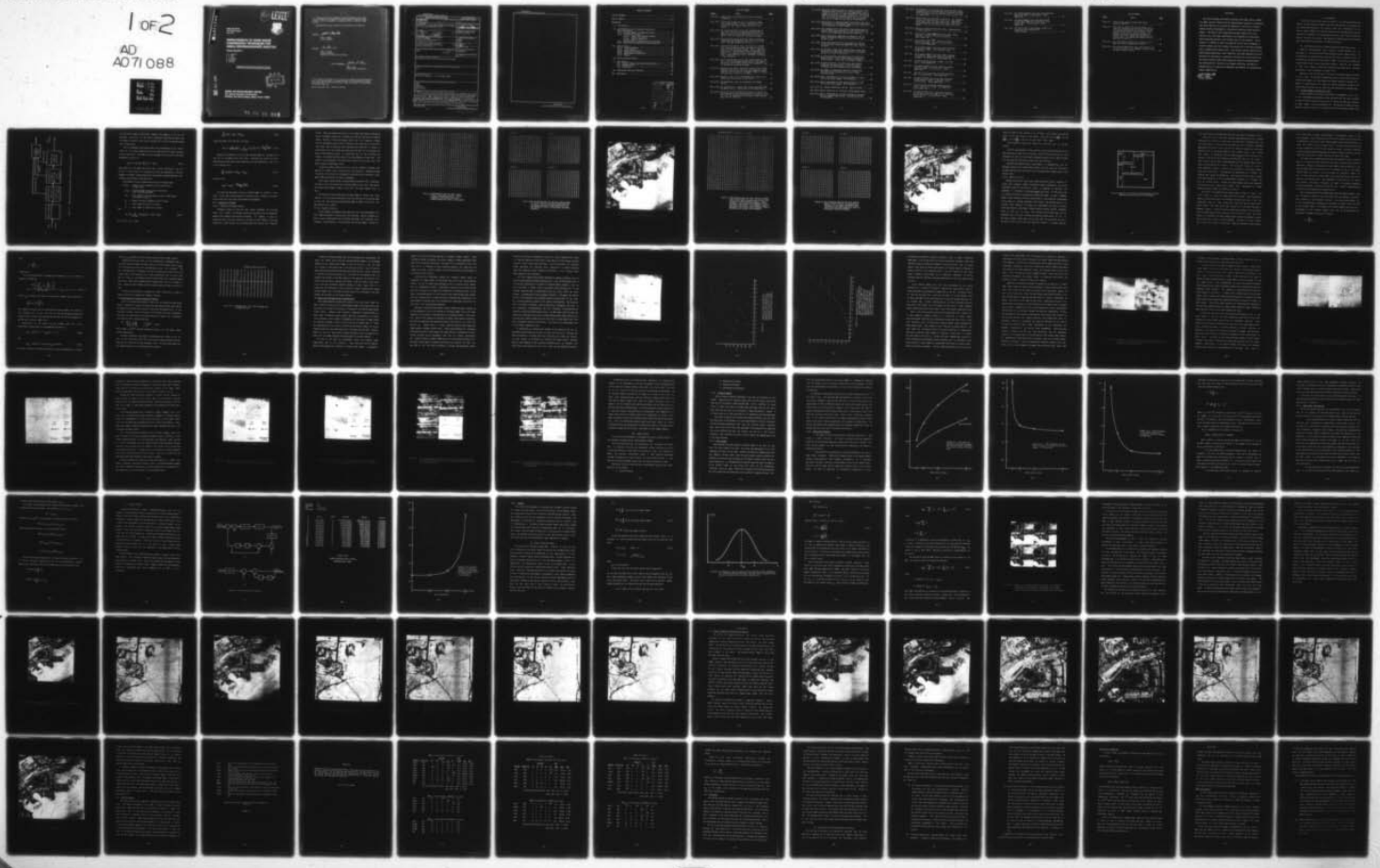
AD-A071 088 PURDUE UNIV LAFAYETTE IND SCHOOL OF ELECTRICAL ENGI--ETC F/G 9/4
IMPROVEMENTS IN SOME IMAGE COMPRESSION TECHNIQUES FOR AERIAL RE--ETC(U)
MAY 79 O R MITCHELL, S C BASS, E J DELP F30602-78-C-0102

UNCLASSIFIED RADC-TR-79-99 NL

1 OF 2
AD A071 088



AD A071088

(12) LEVEL II

RADC-TR-79-99
Final Technical Report
May 1979

IMPROVEMENTS IN SOME IMAGE COMPRESSION TECHNIQUES FOR AERIAL RECONNAISSANCE ANALYSIS

PURDUE UNIVERSITY

O. R. Mitchell
S. C. Bass
E. J. Delp
T. W. Goeddel
A. Tabatabai
P. Stiling

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

DDC
RECEIVED
JUL 12 1979
A

DDC FILE COPY

ROME AIR DEVELOPMENT CENTER
Air Force Systems Command
Griffiss Air Force Base, New York 13441

79 07 12 015

This report has been reviewed by the RADC Information Office (OI) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RADC-TR-79-99 has been reviewed and is approved for publication.

APPROVED: John T. Gamble

JOHN T. GAMBLE
Project Engineer

APPROVED: Fred I. Diamond

FRED I. DIAMOND
Technical Director
Communications and Control Division

FOR THE COMMANDER: John P. Huss

JOHN P. HUSS
Acting Chief, Plans Office

If your address has changed or if you wish to be removed from the RADC mailing list, or if the addressee is no longer employed by your organization, please notify RADC (DCCL) Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.

Do not return this copy. Retain or destroy.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Date Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER RADC-TR-79-99	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) IMPROVEMENTS IN SOME IMAGE COMPRESSION TECHNIQUES FOR AERIAL RECONNAISSANCE ANALYSIS		5. TYPE OF REPORT & PERIOD COVERED Final Technical Report 1 July 78 — 30 September 78
		6. PERFORMING ORG. REPORT NUMBER N/A
7. AUTHOR(s) O. R. Mitchell, T. W. Goeddel S. C. Bass, A. Tabatabai E. J. Delp, P. Stilling		8. CONTRACT OR GRANT NUMBER(s) F30602-78-C-0102
9. PERFORMING ORGANIZATION NAME AND ADDRESS Purdue University School of Electrical Engineering W. Lafayette IN 47907		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 23058P0
11. CONTROLLING OFFICE NAME AND ADDRESS Rome Air Development Center (DCCL) Griffiss AFB NY 13441		12. REPORT DATE May 1979 13. NUMBER OF PAGES 102
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Same		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Same

18. SUPPLEMENTARY NOTES

RADC Project Engineer: Dr. John Gamble (DCCL)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Image
Coding
Compression
Reconnaissance

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this study a number of image coding methods were implemented and applied to aerial reconnaissance photos for relative ranking evaluations by experienced photo-analysts. These coding methods were of three broad types: 1) two-dimensional transform, 2) one-dimensional hybrid, and 3) spatial. All represent significantly improved modifications to techniques studied in a previous work (1). The compression rates used were 1.6 and 0.5 bits/pixel. The effects of channel errors at a 10^{-3} error rate were also simulated and evaluated.

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Date Entered)

TABLE OF CONTENTS

LIST OF FIGURES iv

LIST OF TABLES viii

EVALUATION ix

I. Introduction. 1

II. Two-Dimensional Block Transform Coding with Multi-Class Zones 1
- II.1. Variable Number of Blocks Per Class. 1
- II.2. Selection of Classes 4
  - II.2.1. Energy and Frequency Content. 4
  - II.2.2. Feature Clustering. 5
  - II.2.3. Spatial Criteria. 12
- II.3. Preprocessing to Improve Subjective Ratings. 17
- II.4. Resolution and Quantization Considerations 19

III. Hybrid Coding 36
- III.1. Transform Method. 37
- III.2. Strip Length. 37
- III.3. Quantization Methods. 38
- III.4. DPCM System Configuration 43
- III.5. Summary 49

IV. Block Truncation Coding 49

V. Test Results. 65
- V.1. Original Images and Reconstructed Results 65
- V.2. Ranking Results 73
- V.3. Discussion. 79

VI. Conclusions and Future Research 80

VII. References. 89

Accession For
NTIS GRA&I
DDC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or special
A

LIST OF FIGURES

Figure	Title	Page
Fig. II-1:	Adaptive Multiclass Zone Block Transform Coding Transmitter	2
Fig. II-2:	Classification map for Fig. V-1 using four equal size energy classes. Class 1 represents 16 x 16 blocks with the highest energy and class 4 the lowest.	6
Fig. II-3:	Bit allocation maps for the four classes shown in Fig. II-2. Bits are assigned proportional to the sample variance of each coefficient with the average bit rate 1.5 bits/pixel including overhead.	7
Fig. II-4:	Reconstructed results using four equal size energy classes and the assignment shown in Figs. II-2 and II-3. The mean square error is 35.2.	8
Fig. II-5:	Classification map for Fig. V-1 using four variable size classes based on energy and ratio of low frequency to high frequency energy. Class 1 represents high energy - low frequency; Class 2 represents low energy - low frequency; Class 3 represents high energy - high frequency; Class 4 represents low energy - high frequency.	9
Fig. II-6:	Bit allocation maps for the four classes shown in Fig. II-5. Bits are assigned proportional to the sample variance of each coefficient with the average bit rate 1.5 bits/pixel including overhead	10
Fig. II-7:	Reconstructed results using four variable size classes based on total energy, ratio of low frequency to high frequency energy, and the assignments shown in Figures II-5 and II-6. The m.s.e. is 28.0.	11
Fig. II-8:	Regions in the 16 x 16 transformed block used for determining classification of each block.	13
Fig. II-9:	Two-dimensional zone size histogram with bimodal splits.	18
Fig. II-10:	The rendering of a single pixel target using 90%, 95%, and 98% of the a.c. energy within a transform block	22
Fig. II-11:	Values of cosine transform coefficients, lying on the main diagonal of the transform array at locations (i,i), for a 16 x 16 image block containing a single non-zero pixel of unit intensity	23

Fig. II-12: Recovered intensity value of a single (originally unit intensity) pixel within a 16 x 16 image block. Only a submatrix, of size i x i and including all coefficients whose horizontal and vertical indices are < i, from the full 16 x 16 cosine transform matrix of the isolated pixel block is used for the inverse transformation. . . . 24

Fig. II-13: Quantization of frequency domain coefficients can result in graininess in the decoded image, but more importantly the introduction of false targets 27

Fig. II-14: The justaposition of two inverse transformed blocks of greatly differing graininess is itself an aesthetic defect to which some photo analysts are sensitive 29

Fig. II-15: Use of three-level quantizers on blocks of low frequency information improves the graininess as well as well m.s.e.. 30

Fig. II-16: Three-level quantizers are also beneficial when applied to blocks containing moderately small regions of detail 32

Fig. II-17: On isolated, single pixel target blocks, three-level quantizers can cause the obliteration of the target as well as a rise in the m.s.e. 33

Fig. II-18: At 1.72 bits/pixel, the application of three-level quantization to "city" blocks typically reduces artifact clutter. The m.s.e. can increase, however 34

Fig. II-19: At 0.59 bits/pixel, the three-level quantization scheme may not fare as well. The m.s.e. again increases, and block boundaries are accentuated 35

Fig. III-1: The number of operations required to compute the discrete cosine transform for a strip of data 512 pixels long 39

Fig. III-2: M.S.E. performance of the hybrid systems as a function of the strip length: 1.5 bits/pixel 40

Fig. III-3: M.S.E. performance of the hybrid system as a function of the strip length: 0.5 bits/pixel 41

Fig. III-4: (a) Hybrid transmitter, and (b) hybrid receiver 46

Fig. III-5: Hybrid statistics for a typical reconnaissance image. . . 47

Fig. III-6: M.S.E. performance of the hybrid system in the presence of random channel noise: 16 pixels/strip and 1.5 bits/pixel. 48

Fig. IV-1: In designing a one bit quantizer given the data values (assumed to be a continuous density function), one must find a threshold value X_{th} and two output levels a and b, respectively 51

Fig. IV-2: Results using various fidelity criteria. All representations are 2.0 bits/pixel. Upper left: MSE; Upper right: MAE; Lower left: MP; Lower right: MSE and also assuming image data uniformly distributed with each block 54

Fig. IV-3: Results of Hybrid Formulation of BTC. Representation is 1.88 bits/pixel. 58

Fig. V-1: The original image "SAM3" used in the photo analyst rankings. Size 512 x 512 at 8 bits/pixel, Photo number 500. 59

Fig. V-2: The original image "AP2" used in the photo analyst rankings. Size 512 x 512 at 8 bits/pixel, Photo number 600 60

Fig. V-3: The frequency-domain four class zonal technique as applied to SAM3: 1.6 bits/pixel, Photo number 549. 61

Fig. V-4: The frequency-domain four class zonal technique (with spatial domain preprocessing) as applied to SAM3 - 1.6 bits/pixel. Photo number 562 62

Fig. V-5: The hybrid method applied to SAM3: 1.6 bits/pixel. Photo number 514. 63

Fig. V-6: The moment-preserving block truncation method applied to SAM3: 1.6 bits/pixel. Photo number 501. 64

Fig. V-7: The basic block truncation algorithm applied to AP2: 1.6 bits/pixel. Photo numer 631. 66

Fig. V-8: The spatial-domain four class zonal technique (using unequal classes) as applied to AP2: 1.6 bits/pixel. Photo number 643 67

Fig. V-9: The basic block truncation method applied to a 256 x 256 section of AP2: 1.6 bits/pixel. Photo number 407. 68

Fig. V-10: The spatial-domain four class zonal technique (using equal classes) as applied to a 256 x 256 section of AP2: 1.6 bits/pixel. Photo number 405. 69

Fig. V-11: The frequency-domain four class zonal method as applied to SAM3: 0.5 bits/pixel. Photo number 538. 70

Fig. V-12: The spatial-domain four class zonal method as applied to SAM3 with a channel bit error rate of 10^{-3}: 1.6 bits/pixel. Photo number 523. 71

Fig. V-13: The hybrid coder in application to AP2 with a channel bit error rate of 10^{-3}: 1.6 bits/pixel. 72

LIST OF TABLES

Table	Title	Page
Table V-1:	Abbreviations used in ranking results presented in Table V-2.	74
Table V-2:	"Raw data" as taken from analysts' forms. Also shown for each method are an average ranking and the computed mean-square error (m.s.e.) and mean absolute error (m.a.e.) for that method. The average ranking was calculated by dropping the highest and lowest rankings and averaging the remaining three. The method name abbreviations are described in Table V-1	75
Table VI-1:	The instantaneous coding rates available to the MAPS algorithm as implemented at Purdue.	86
Table VI-2:	Inserting new MAPS coding rate levels (marked with an asterisk) results in a somewhat finer partition for better adaptive coding	87

EVALUATION

This effort extended the efforts performed under RADC contract F30602-75-C-0082, entitled "Coding of Aerial Reconnaissance Images for Transmission Over Noisy Channels" by an objective comparison of the effect of source/channel encoding improvements in transmitting digitized, high resolution imagery. The bases of this comparison were mean squared error, mean absolute error, and, most significantly, the assessment of professional photo-interpreter/analysts (the ultimate users of such imagery). Source encoding is intended to reduce non-essential (to the user) redundancy in original imagery, but this increases the sensitivity of the source-encoded data to communication channel noise. Thus channel encoding (efficient bit apportionment/quantization, data formatting, and added redundancy for error detection and correction) is necessary to minimize this noise sensitivity. The present research effort which emphasized improved bit apportionment and quantization for realistic noisy channel conditions, provides for enhanced quality in reducing the bandwidth requirements for reconnaissance imagery communication.

John T. Gamble

JOHN T. GAMBLE
Project Engineer

I. Introduction

An earlier study [1] has shown the potential for non-entropy-preserving coding of high resolution imagery such that photo-analysts are pleased with the reconstructed pictures. This present study has concentrated on the source coding methods with the most potential for this application and has resulted in refinement of three methods (multiclass zone block transform, hybrid, and block truncation). New subjective rankings have been performed which show significant advances in coding performance.

II. Two-Dimensional Block Transform Coding with Multiclass Zones

Adaptive two-dimensional block transform coding methods such as described by Chen and Smith [2] have produced the best ranked results. In such methods, the Two-Dimensional Fast Discrete Cosine Transform [3,4] is performed over sub-blocks of the original image. The blocks are sorted into classes based on one or more statistical features. Then bits are assigned to each class based on coefficient variances within each class. Detailed description of such techniques are found in [1].

Reported in this section are (1) the use of non-equal numbers of blocks in each class; (2) methods of assigning blocks to classes which result in improved mean-square error performance and higher subjective ratings; (3) a method of preprocessing the image to improve subjective ratings of the reconstructed pictures; and (4) resolution and quantization considerations.

II.1. Variable Number of Blocks Per Class

A typical adaptive multiclass zone method is diagrammed in Fig. II-1. In the implementation described by Chen and Smith [2], the total ac energy is used to classify each block into one of 4 classes, so that each class has an equal number of blocks assigned to it. The purpose of having an equal number of blocks per class is to insure easily that the average coding rate

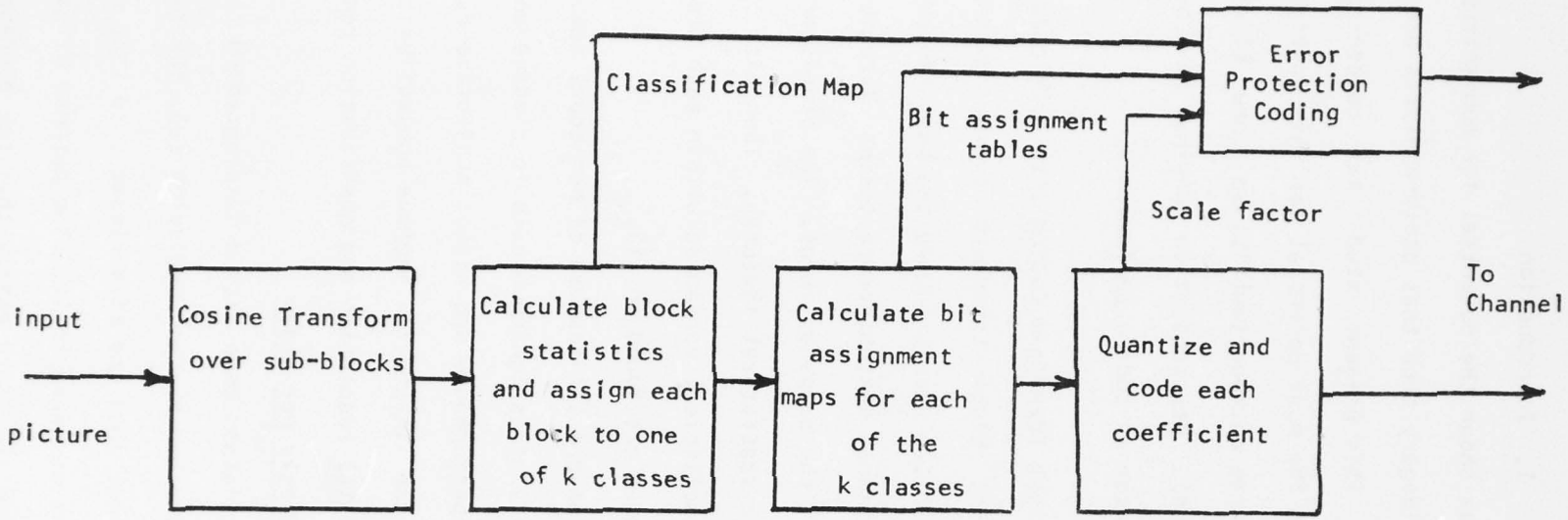


Fig. II-1. Adaptive Multiclass Zone Block Transform Coding Transmitter

over the entire image is maintained. However, this appears to be an unnecessary restriction in that the bit assignment algorithm can easily use the number of blocks in each class to assign bits so that the average coding rate is maintained.

The bit assignment algorithm assigns bits simultaneously to all coefficients in all classes proportional to the logarithm of the sample variance of each coefficient. The number of bits assigned to the (u,v)th coefficient belonging to class k is

$$N_k(u,v) = \frac{1}{2} \log_2 [\sigma_k^2(u,v)] - \log_2 D \qquad \text{(II-1)}$$

where $\sigma_k^2(u,v)$ is the sample variance of the (u,v)th coefficient over all blocks in class k and D is a constant which sets the compression. This assignment is optimal in mean-square-error performance assuming Gaussian distributed, uncorrelated coefficients [5].

The parameter D can be determined using the following steps:

Let B_{dc} = number of bits assigned to the dc coefficient in each class.

B_{avg} = desired average coding rate in bits/pixel (not including overhead).

B_k = total number of bits assigned to the ac coefficients of a block in class k.

C_k = number of blocks assigned to the k^{th} class.

N = total number of pixels in a block

M = total number of pixels in a picture.

Then

$$B_k = \frac{1}{2} \sum_{(u,v)\neq 0} \log_2 \sigma_k^2(u,v) - (N-1) \log_2 D \qquad \text{(II-2)}$$

for k = 1,2,...,K. Also,

$$\sum_{k=1}^{K} C_k(B_k + B_{dc}) = M \cdot B_{avg} \qquad \text{(II-3)}$$

substituting EQ. II-2 into EQ. II-3 gives

$$\log_2 D = \frac{N}{2M(n-1)} \sum_k \{C_k \sum_{(u,v)\neq 0} \log_2 \sigma_k^2(u,v)\} + \frac{B_{dc} - N \cdot B_{avg}}{N-1} \qquad \text{(II-4)}$$

Because of round-off error, the exact desired number of assigned bits may not be assigned on the first trial. Therefore, the value of D can be modified as follows where primes indicate the values obtained on the first trial: From Eq. II-3

$$\sum_{k=1}^{K} C_k(B_k - B_k') = M(B_{avg} - B_{avg}') \qquad \text{(II-5)}$$

using Eq. II-2

$$\log_2 D = \log_2 D' - \frac{N(B_{avg} - B_{avg}')}{N-1} \qquad \text{(II-6)}$$

Of course the advantage of using a variable number of blocks in each class is that small unique regions of a picture (e.g., islands in an ocean) can be given their own class and optimum bit assignment.

II.2. Selection of Classes

II.2.1. Energy and Frequency Content

Although ac energy is often the most useful parameter for dividing blocks into classes (low energy classes get few bits), such an assignment neglects the frequency distribution of energy. For example, a group of blocks may have a low total energy but with most of their energy in the high frequencies. Other blocks with low energy might be primarily low frequency

blocks. These two groups would be put in the same class based on energy but the bit assignment would have to spread over both the high and low frequency ranges. However, if these two groups could be put into separate classes, the bit assignment would be quite different, thus resulting in a more efficient code. Such a system has been implemented by using the ac energy to divide blocks into two classes and then using the ratio of low frequency energy to high frequency energy to sub-divide each of the resulting two classes. The sample mean was used as the class boundary in each case. This system is to be compared to one which uses energy only to divide into four equal classes as described in [2].

Using the original shown in Fig. V-1, the two methods discussed were applied to obtain a total coding rate of 1.5 bits/pixel. Overhead allowed (including error protection) averaged 15.5 bits/block thus leaving an average of 368.5 bits/block for the bit assignment.

The 4 equal classes based on energy method resulted in the classification map of Fig. II-2 and the bit allocation maps of Fig. II-3. The resulting reconstructed image is shown in Fig. II-4. The mean square error is 35.2.

The energy and ratio method with variable number of blocks per class resulted in the classification map of Fig. II-5 and the bit allocation maps of Fig. II-6. The resulting reconstructed image is shown in Fig. II-7. The mean square error is 28.0.

II.2.2. Feature Clustering

The performance improvement shown above has led to the development of a more adaptive method of choosing the class features. Several frequency regions of each transformed block are defined including low frequency, mid-frequency, high-frequency, and horizontal and vertical edges. Typical re-

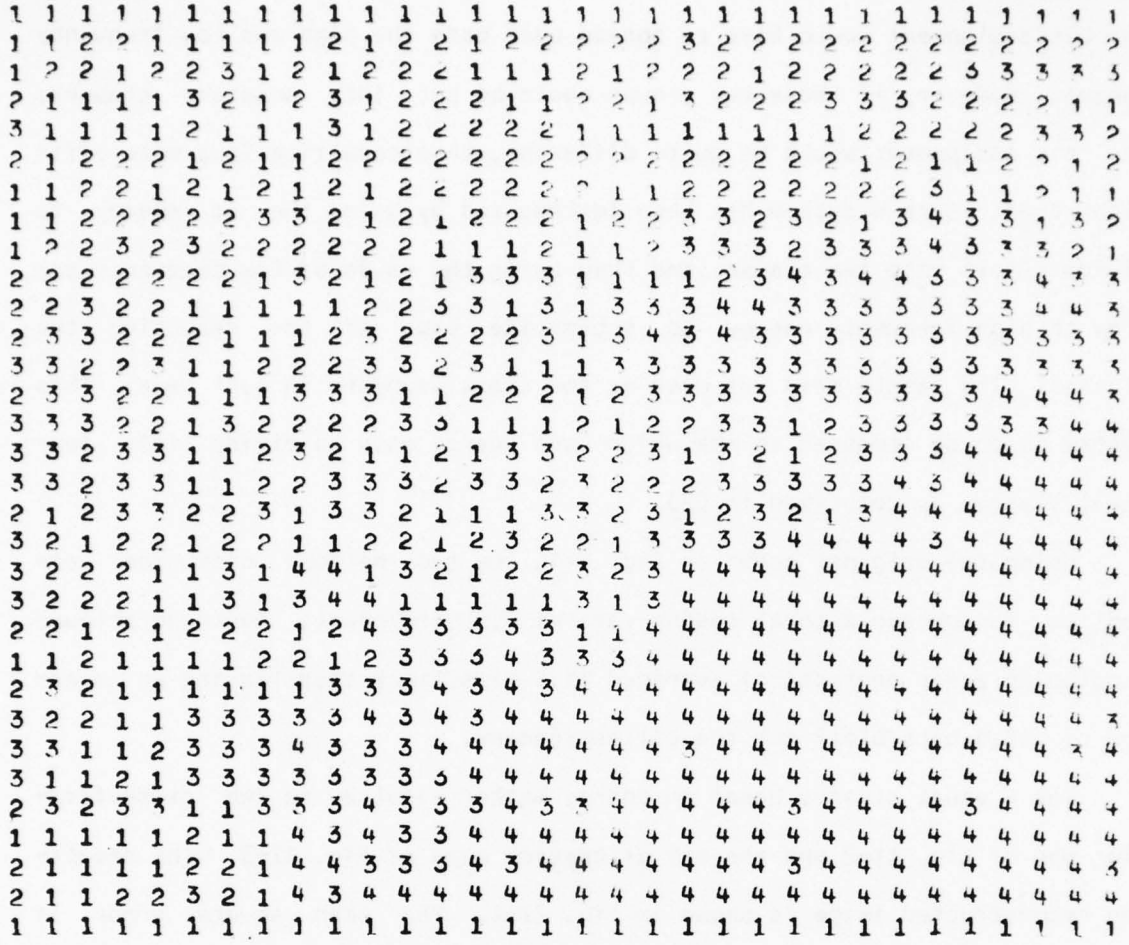


Fig.II-2. Classification map for Fig.V-1 using 4 equal size energy classes. Class 1 represents 16x16 blocks with the highest energy and Class 4, the lowest.

CLASS 1

8	6	5	4	4	3	3	3	2	2	2	2	2	2	1	1
7	5	5	4	4	3	3	3	2	2	2	2	2	1	1	1
6	5	4	4	4	3	3	3	2	2	2	2	2	1	1	1
5	4	4	4	4	3	3	3	2	2	2	2	1	1	1	1
5	4	4	4	4	3	3	3	2	2	2	2	2	1	1	1
4	4	4	3	3	3	3	2	2	2	2	2	1	1	1	1
5	4	4	3	3	3	3	2	2	2	2	1	1	1	1	1
5	3	3	3	3	3	2	2	2	2	2	1	1	1	1	1
3	3	3	3	2	2	2	2	2	2	2	1	1	1	1	1
4	3	2	2	2	2	2	2	2	2	2	1	1	1	1	1
4	2	2	2	2	2	2	2	2	1	2	1	1	1	1	1
3	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1
4	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1
3	2	2	2	2	1	2	1	1	1	1	1	1	1	1	1
3	2	2	2	2	1	1	1	1	1	1	1	1	1	1	0
4	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1

CLASS 3

6	5	4	3	3	2	2	2	2	2	1	1	1	1	1	1
5	4	3	3	3	3	2	2	1	1	1	1	1	1	1	1
4	3	3	3	2	2	2	2	1	1	1	1	1	1	0	0
3	3	3	3	3	2	2	2	2	1	1	1	1	1	0	1
3	3	3	3	3	2	2	2	2	1	1	1	1	1	1	0
3	3	3	3	2	2	2	2	1	1	1	1	1	1	0	0
3	2	3	2	2	2	1	2	1	1	1	1	1	1	1	0
3	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1
2	2	2	2	2	2	1	1	1	1	1	1	1	0	1	0
2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	0
2	2	1	1	1	1	1	1	1	1	1	1	1	0	1	0
2	2	2	1	1	1	1	1	1	1	1	1	0	0	0	0
2	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0
1	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0
1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0

CLASS 2

8	5	5	4	4	4	3	3	3	2	2	2	2	2	2	1
5	5	4	4	4	3	3	3	3	2	2	2	2	2	2	1
5	4	4	4	4	3	3	3	2	2	2	2	2	2	1	2
4	4	4	4	4	3	3	3	2	2	2	2	2	2	1	2
4	4	4	4	3	3	3	3	2	2	2	2	2	2	1	1
4	4	4	4	3	3	3	2	2	2	2	2	2	1	1	1
4	3	3	3	3	3	3	2	2	2	2	2	2	1	1	1
3	3	3	3	3	3	2	2	2	2	2	2	1	2	1	1
3	3	3	3	3	2	2	2	2	2	2	2	2	2	1	1
3	3	3	3	2	3	2	2	2	2	2	1	1	1	1	1
3	3	3	2	2	2	2	2	2	2	2	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1
2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	1	2	1	1	1	1	1	1

CLASS 4

8	3	2	1	1	0	0	0	0	0	0	0	0	0	0	0
3	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0
2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig.II-3 Bit allocation maps for the four classes shown in Fig.II-2.Bits are assigned proportional to the sample variance of each coefficient with the average bit rate 1.5 bits/pixel including overhead.



Fig. II-4. Reconstructed results using four equal size energy classes and the assignment shown in Figs. II-2 and II-3. The mean square error is 35.2.

CLASSIFICATION MAP FOR EACH BLOCK

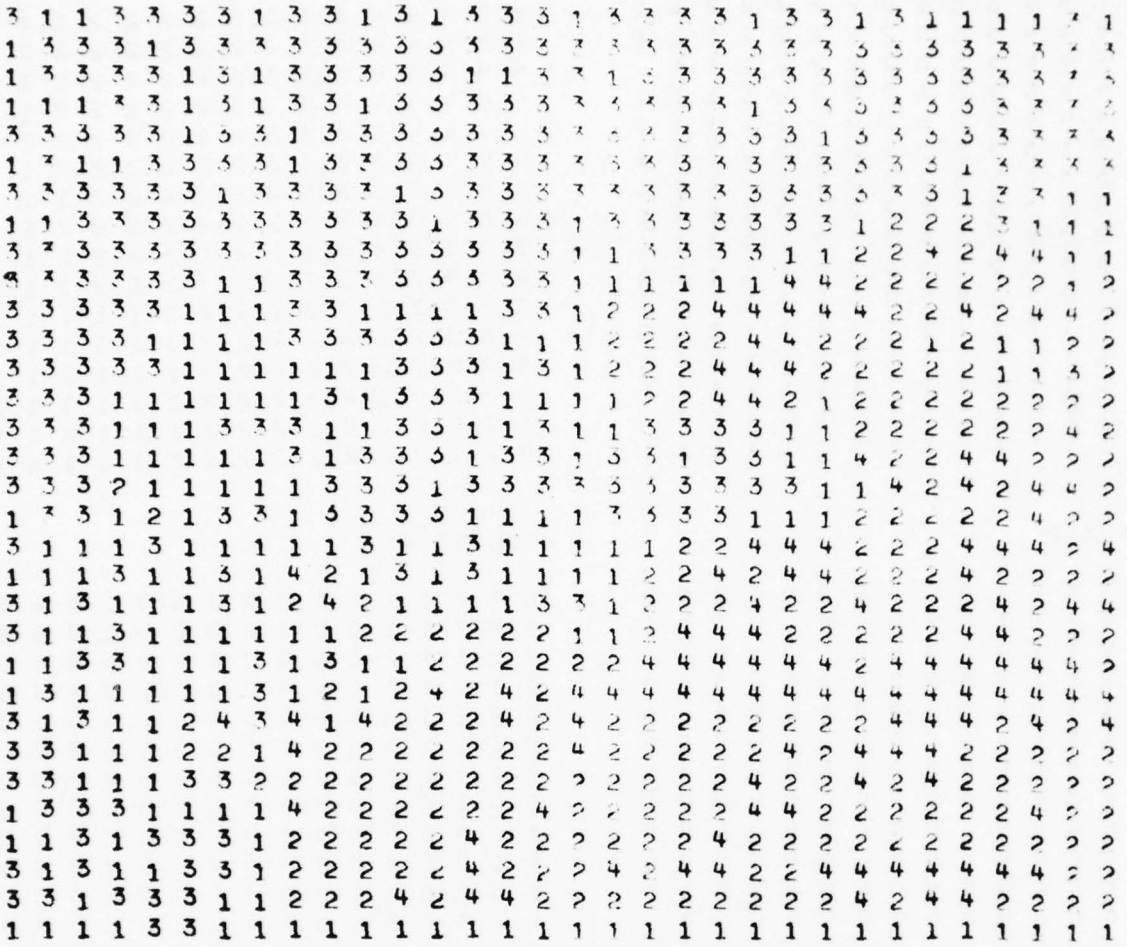


Fig.II-5 Classification map for Fig.II-1 using 4 variable size classes based on energy and ratio of low frequency to high frequency energy. Class 1 represent, high energy, low frequency; Class 2 represents low energy, low frequency; Class 3 represents high energy, high frequency; Class 4 represents low energy, high frequency.

CLASS 1

8	6	5	5	4	3	3	3	3	2	2	2	2	2	1	1
7	5	5	4	4	3	3	3	2	2	2	2	2	1	1	1
6	5	4	4	4	3	3	3	2	2	2	2	1	1	1	1
5	4	4	4	4	3	3	3	2	2	2	2	1	1	1	1
5	4	4	4	4	3	3	3	2	2	2	1	1	1	1	1
4	3	3	3	3	3	3	2	2	2	2	2	1	1	1	1
4	3	3	3	3	3	3	2	2	2	2	1	2	1	1	1
4	3	3	3	2	3	2	2	2	2	2	2	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1
4	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1
4	2	2	2	2	2	2	1	2	1	2	1	1	1	1	1
3	2	2	1	2	1	2	1	1	1	1	1	1	1	1	1
4	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

CLASS 3

8	5	4	4	4	3	3	3	3	2	2	2	2	2	2	1
6	5	4	4	4	3	3	3	2	2	2	2	2	2	2	2
5	4	4	4	4	3	3	3	2	2	2	2	2	2	2	2
5	4	4	4	4	3	3	3	3	2	2	2	2	2	1	1
5	4	4	4	3	3	3	3	2	2	2	2	2	2	2	1
4	4	4	4	3	3	3	3	2	2	2	2	2	2	1	1
4	4	4	3	3	3	3	2	2	2	2	2	2	1	1	1
4	3	3	3	3	3	2	2	2	2	2	2	2	2	1	1
3	3	3	3	3	3	2	2	2	2	2	2	2	2	1	1
4	3	3	3	3	3	2	2	2	2	2	2	2	1	1	1
3	3	3	3	2	2	2	2	2	2	2	1	1	1	1	1
3	3	3	2	2	2	2	2	2	2	1	1	1	1	1	1
3	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1
3	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1
3	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1

CLASS 2

8	4	3	2	2	1	0	0	0	0	0	0	0	0	0	0
4	3	2	2	1	1	0	0	0	0	0	0	0	0	0	0
3	2	2	1	1	0	0	0	0	0	0	0	0	0	0	0
2	2	1	1	1	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

CLASS 4

8	3	2	1	1	0	0	0	0	0	0	0	0	0	0	0
3	2	2	1	1	0	0	0	0	0	0	0	0	0	0	0
2	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0
2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
2	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0
2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure II-6 Bit allocation maps for the four classes shown in Fig.II-5. Bits are assigned proportional to the sample variance of each coefficient with the average bit rate 1.5 bits/pixel including overhead.

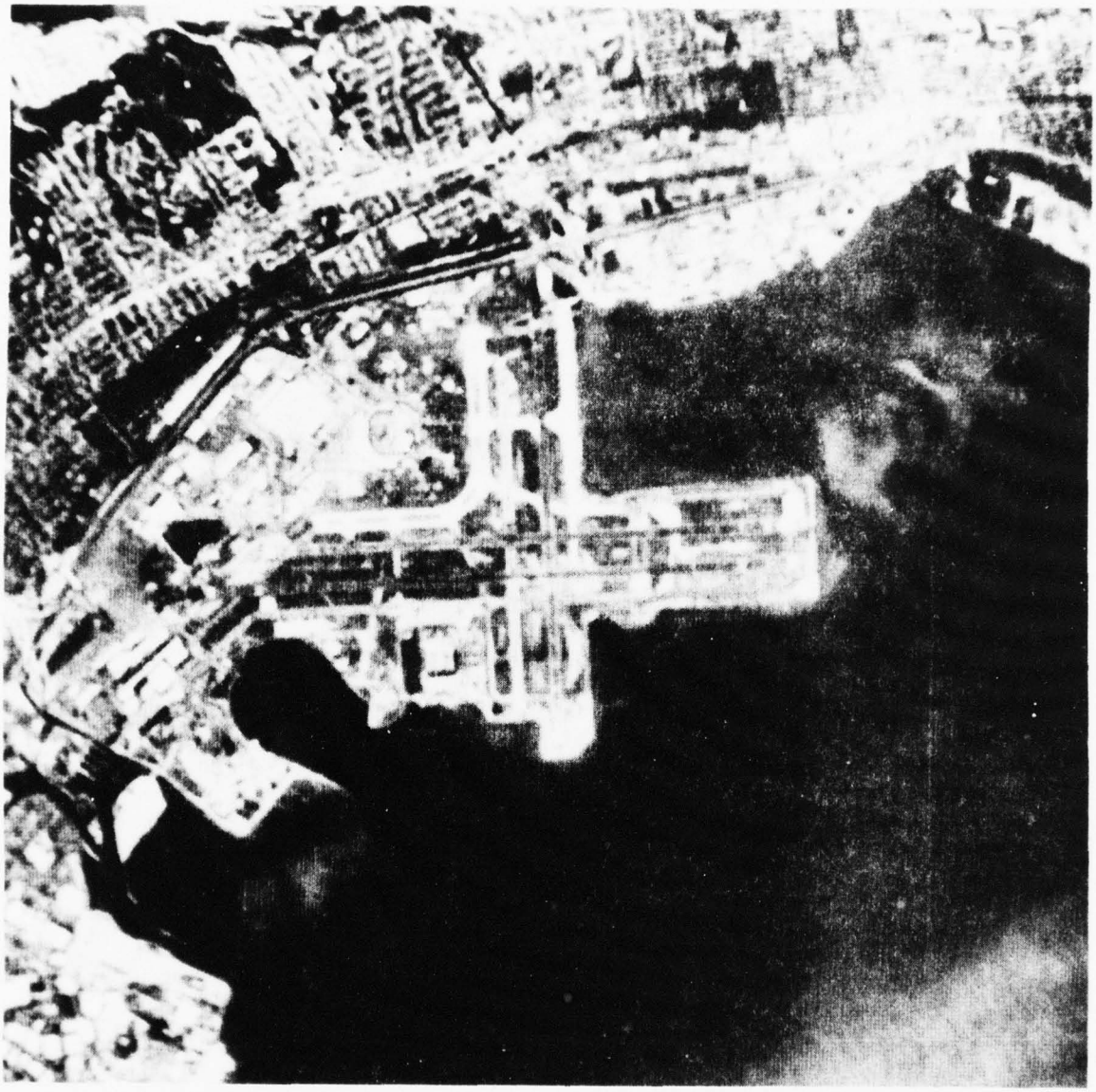


Fig. II-7. Reconstructed results using four variable size classes based on total energy, ratio of low frequency to high frequency energy, and the assignments shown in Figures II-5 and II-6. The m.s.e. is 28.0

gions are shown in Fig. II-8 for a 16 x 16 block. Four features are defined based on the total energy in each region. These are: (1) $\log \frac{LOW}{MED}$; (2) $\log \frac{LOW}{HIGH}$; (3) $\log \frac{MED}{EDGE}$: and (4) log (LOW + MED + HIGH + EDGE).

The histograms of these four features are collected over all picture blocks.

A clustering procedure is then used to find the histogram which can be most obviously divided into two classes. After the first division, all four histograms for each class are again examined and each class is again divided into two using the most useful feature for each division.

Although this method offers a large amount of adaptability, the improvement gained is not significantly better than that using ratio and energy (described earlier) for typical aerial reconnaissance imagery used in the present study.

II.2.3. Spatial Criteria

Another of the four class zonal coding techniques we have studied involves both frequency domain and spatial domain information. The cosine transform of each 16 x 16 block of pixels is examined to determine the bandwidth of information contained within that block. In particular, the a.c. energy contained in successively larger and larger circular frequency domain zones (centered at the dc coefficient) is calculated as a percentage of the total a.c. energy contained in this block. The minimum radius of a circular zone required to contain at least 90% of this total ac energy within the block is calculated. "Water" blocks typically exhibit bandwidth radii of from four to seven. "City" blocks, on the other hand, take on bandwidth radii up to and including the maximum allowed amount of 15. In the spatial domain, these same circular zone radii are calculated but now on the basis of the circular zone size required to obtain a maximum absolute

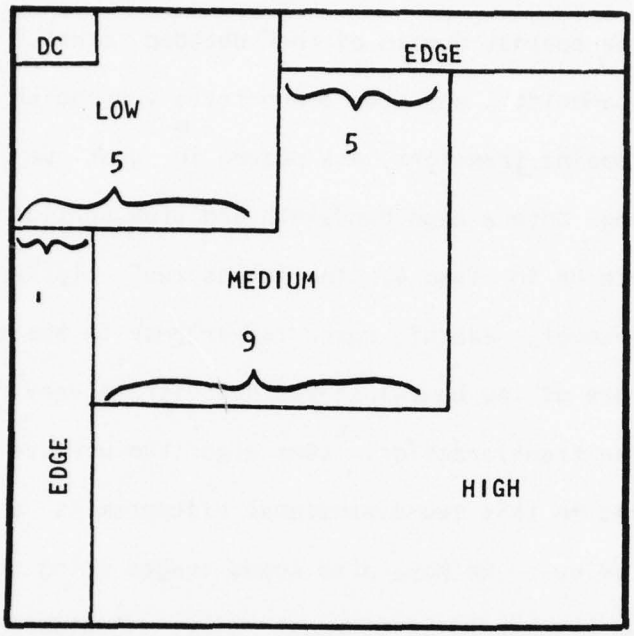


Fig.14-8 Regions in the 16x16 Transformed Block Used for Determining Classification of Each Block.

error upon inverse transformation that does not exceed a threshold of 12.5.

These two circular zone radii (based on 90% bandwidth and 12.5 maximum absolute error) allow us to construct a two dimensional histogram that lists the total number of picture blocks that have any given possible bandwidth and require any given circular zone size to achieve a moderately low maximum absolute error in the spatial domain of the decoded block. Those blocks that exhibit low bandwidth, and also demonstrate very quick pointwise convergence using the cosine transform, are placed in what we call zone 1. Those blocks having both a high bandwidth and slow spatial domain convergence, are defined to be in class 4. The "class two" blocks are also of wide bandwidth, however, exhibit quick convergence in the spatial domain. Class three blocks are of low bandwidth yet demonstrate very stubborn convergence upon inverse transformation. (One algorithm we have used to establish class boundaries in this two-dimensional histogram is a bi-modal one that we discuss below. We have also coded images using class boundaries that forced all four classes to be of equal size. Evaluations of images coded using both schemes are reported on below.)

The result is an algorithm that places "water" blocks in class 1 and "city" blocks in class 4, respectivey, very reliably. More importantly, however, we have found that this technique isolates into class 3 two very troublesome types of image blocks; the "coastline" and the "boat-in-the-water" blocks. The former occurs when a group of 16 x 16 pixels overlaps a boundary between a region of high ac energy and a region of low ac energy. Since the high energy pixels occupy only a portion of the total 256 pixel region, it is possible for other categorization schemes erroneously to place such a block into a lower ac energy category. The result is a poor rendering of the "city" portion of the block. "Boat-in-the-water" blocks exhibit

one or two pixels of highly isolated detail in a background field of very low bandwidth. Again observing only the total ac energy with the block (or its bandwidth) could cause a more simple-minded categorization technique to place this block in, say, a "water" class. The result would be a blurring or even a complete loss of the small potential target. By paying attention to both frequency bandwidth and pointwise spatial domain covergence, we have found that we may very reliably flag both of these block types.

Of course such an approach does require the repeated inverse transformation of larger and larger subsets of the cosine transform coefficients of each block. However, by making use of the normally high correlation of the spatial domain radius from one block to its neighbor, we may cut this computational labor to a minimum. The initial guess for the spatial domain radius of each new block is simply taken as the resulting radius value calculated for the previous block. Then adjustments in the zone size to obtain the necessary point-wise convergence are made in an upward or downward direction as needed. Typically only three or four inverse transformations of each block are needed during categorization.

Once the above-mentioned two-dimensional spatial-frequency domain zone size histogram is formed, it must be divided into four regions. This division is based on a "most bimodal" criterion. If we define each element h_{ij} of the two-dimensional histogram $\underline{H}$ as the number of blocks in the (i,j) bin of the histogram (where i corresponds to the spatial domain zone size and j corresponds to the frequency domain zone size) we can define two one-dimensional histograms $\underline{x}$ and $\underline{y}$ with elements

$$x_i = \sum_{j=1}^{16} h_{ij}$$

and

$$y_j = \sum_{i=1}^{16} h_{ij}$$

respectively.

For any one-dimensional histogram $\underline{a}$ with elements a_i we can define a measure of bimodality

$$b_i = \frac{\min \left[\sum_{j=i_{min}}^{i-1} a_i, \sum_{k=i+1}^{16} a_i \right]}{\max \{a_i, 0.5\}}, \quad i = i_{min} + 1, \cdots, 15 \qquad \text{(II-7)}$$

where i_{min} is chosen to be equal to the smallest integer value j such that

$$\sum_{i=1}^{j} a_j \geq .25 \sum_{k=1}^{16} a_k.$$

This condition ensures that at least 25% of the total number of blocks in the histogram will be put into one of the two low energy classes and prevents these classes from receiving too few bits in the bit assignment algorithm due to a small sample size.

Using (II-7), we can define the most bimodal point for a one-dimensional histogram and its corresponding location as

$$b_{max} = \max_i \{b_i\}, \quad i = i_{min}+1, \cdots, 15 \qquad \text{(II-8)}$$

and

$$i_{max} = \text{index of } b_i \text{ at which } b_{max} \text{ occurred.} \qquad \text{(II-9)}$$

The actual location of the most bimodal split will be immediately on either

side of i_{max}, whichever division yields the most nearly equal classes.

Applying (II-8) and (II-9) to the one-dimensional histograms $\underline{x}$ and $\underline{y}$, we choose the most bimodal of these (the one with the largest value of b_{max}) for the primary division of the two-dimensional zone size histogram. Two new one-dimensional histograms may now be formed, one on each side of the primary split, with the major axis in both cases being the criterion that was not used in the primary division (either spatial or frequency domain zone size). These new one-dimensional histograms may then be divided at their respective most bimodal points by again applying (II-8) and (II-9) to each.

A typical two-dimensional histogram for a 256 x 256 picture is shown in Figure II-9 with its appropriate bimodal divisions.

II.3 Preprocessing to Improve Subjective Ratings

Since the ultimate performance criterion for our coding was human photo analyst subjective ratings, we found that some modifications could improve the subjective ranking while actually raising the mean-square-error in the resulting picture. Such is the case of a non-linear gray level transformation prior to coding and the inverse process after coding. The transformation used for 8-bit original data was:

$$\begin{array}{lll} \text{for} & 245 < x \leq 255 & y = x \\ & 23 \leq x \leq 245 & y = 40x^{0.4} - 116.17 \\ & 0 \leq x < 23 & y = x \end{array}$$

Such a power law ($X^{0.4}$) has been proposed as similar to the human visual system response [6].

The more traditional logarithmic transformation was found to be too flat at the high gray levels (the inverse after coding reconstruction emphasized the ringing due to high frequency loss). This particular power law was judged subjectively most pleasing by the authors.

Frequency Domain Zone Size

0	0	0	0	2	3	3	6	1	7	2	0	0	0	0	0
0	3	6	6	2	0	5	1	3	0	0	0	0	0	0	0
0	0	2	1	4	4	2	1	0	0	0	0	0	0	0	0
0	0	3	0	2	0	1	0	0	0	0	0	0	0	0	0
0	0	0	2	4	1	0	0	3	0	0	0	0	0	0	0
0	0	2	0	0	1	1	0	0	0	0	0	0	0	0	0
0	0	1	0	2	1	0	0	0	0	0	0	0	0	0	0
0	0	1	1	2	2	1	3	1	0	0	0	0	0	0	0
0	0	4	1	3	4	7	5	3	0	0	0	0	0	0	0
0	0	3	3	13	9	14	7	1	0	0	0	0	0	0	0
0	0	2	2	11	3	7	6	3	2	0	0	0	0	0	0
0	0	0	1	1	3	3	9	12	6	0	0	0	0	0	0
0	0	0	0	2	0	0	3	1	1	1	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig. II-9. Two-dimensional zone size histogram with bimodal splits.

Subjective rankings showed that, on the average, the preprocessed imagery was ranked better than the non-preprocessed imagery. As an example compare the two reconstructed images in Figs. V-3 and V-4. Fig. V-3 uses the 4-feature method described in the previous section. Fig. V-4 used the same method except preprocessing precedes coding and the inverse operation is applied following reconstruction. The effect of the preprocessing is to emphasize the energy in the dark regions of the image so that more bits are used to code the darker regions than normal. Since the human is more sensitive to the same incremental additive changes in a dark region than in a bright one, the coding more clearly matches the human's response even though the mean square error increased (the mean square error decreases in the dark regions and increases in the bright regions).

II.4 Resolution and Quantization Considerations

Most image bandwidth compression studies in the past have aimed for good aesthetic reception (of a coded-decoded image) by the human brain, and have often relied upon the existence of substantial correlation among neighboring pixels. However, when aesthetic appearance is made secondary to sheer useful information content, and/or when each isolated pixel is potentially important, then a number of judgements and techniques in image coding are required to change. We will use this section to examine some of these new considerations at bit rates in the 0.5-2.0 bits/pixel range. All of our example blocks (of size 16x16 pixels) will be drawn from the "AP2" original shown in Fig. V-1. Over several regions (especially the city areas at the upper left) the image bandwidth here approaches the spatial Nyquist.

Our work for the Rome Air Development Center has required image compressions down to 0.5 bits/pixel. Since most (but not all) spatial domain coding methods are limited to 1 bit/pixel and above, a substantial

amount of our effort has been applied to frequency domain coders. These transform coding techniques are easily capable of image compression rates down to 0.5 bits/pixel and below. Thus most of the discussion of this section will be in reference to zonal transform coding of, for simplicity, the single class type. Here the image is artificially broken into sub-images of size 16x16 called "blocks".

A particularly troublesome problem for frequency domain coders is presented by image regions of high detail that are of very limited spatial extent. In Fig. V-1 these could be boats of one or two pixel width appearing on a background of water. Alternately a narrow strip of coastline may turn out to occupy only an edge of an otherwise low frequency block. In any case, and depending upon the intricacies of the coding method used, rendering such isolated details without uniformly increasing the transmission bit rate presents a problem that, while aesthetically negligible, may well be of real concern for reconnaissance purposes.

One means of handling such special cases, without uniformly easing up on the compression rate, is to provide an increased number of bits for these (hopefully few) blocks of isolated detail on an adaptive basis. This naturally implies that schemes are available to detect these special situations. One detection approach that seems particularly attractive is to total the percent a.c. energy within a block, working from the lower toward the higher spatial frequency coefficients. (These sums-of-squares of transform coefficients would be calculated by the coder during its evaluation of coefficient variances for bit assignment. Thus the a.c. energy calculations will require further arithmetic additions but no new multiplications.) If a noticeably large number of squared coefficients are required to total to, say, 90% of the full block transform a.c. energy, then additional coding

bits for this block are apparently called for. Such an approach will clearly flag any image block containing a large amount of high frequency energy, such as a "city" region mentioned above. Since these blocks are obviously prime candidates for receiving extra coding bits -- to better represent their more important higher frequency coefficients -- this "energy threshold" approach looks promising.

For "city" image blocks widely dominated by regions of detail, this energy threshold detection method performs well. Fig. II-10 indicates that it is less than satisfactory for blocks of isolated detail, however. Here we have extracted a single "boat in the water" block from Fig. V-1. In the original image the boat is represented by only one pixel while the water presents a background of slowly changing intensity. The upper left quarter of Fig. II-10 represents the recovered image block where 90% of the total a.c. transform energy is used for the inverse transformation. The one-pixel boat is no where in evidence. Even at 95% energy threshold, the boat is only marginally distinguishable and thus vulnerable to obliteration by transform coefficient quantization errors. At 98% (lower left) the boat has become quite distinct, but at such a high threshold most other blocks within the scene were found to have their bit rates set at a much higher level then their information content could justify. The result is an unacceptable loss in the overall compression rate.

To understand the isolated pixel problem in the context of the two dimensional Cosine Transform, we refer to Figs. II-11 and II-12. We apply the Cosine Transform to an image test block consisting of a single unit intensity pixel placed in the middle of a field of zero (dark) pixels. Scanning down the main diagonal of the resulting transform array, we encounter the coefficient values plotted in Fig. II-11. While the two dimensional Fourier

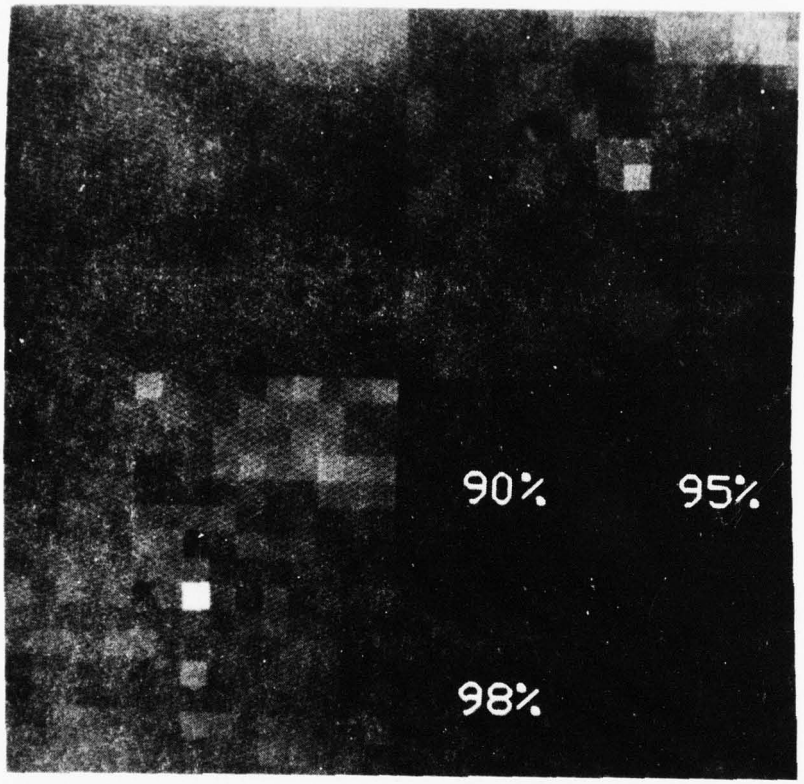


Fig. II-10 The rendering of a single pixel target using 90%, 95%, and 98% of the a.c. energy within a transform block.

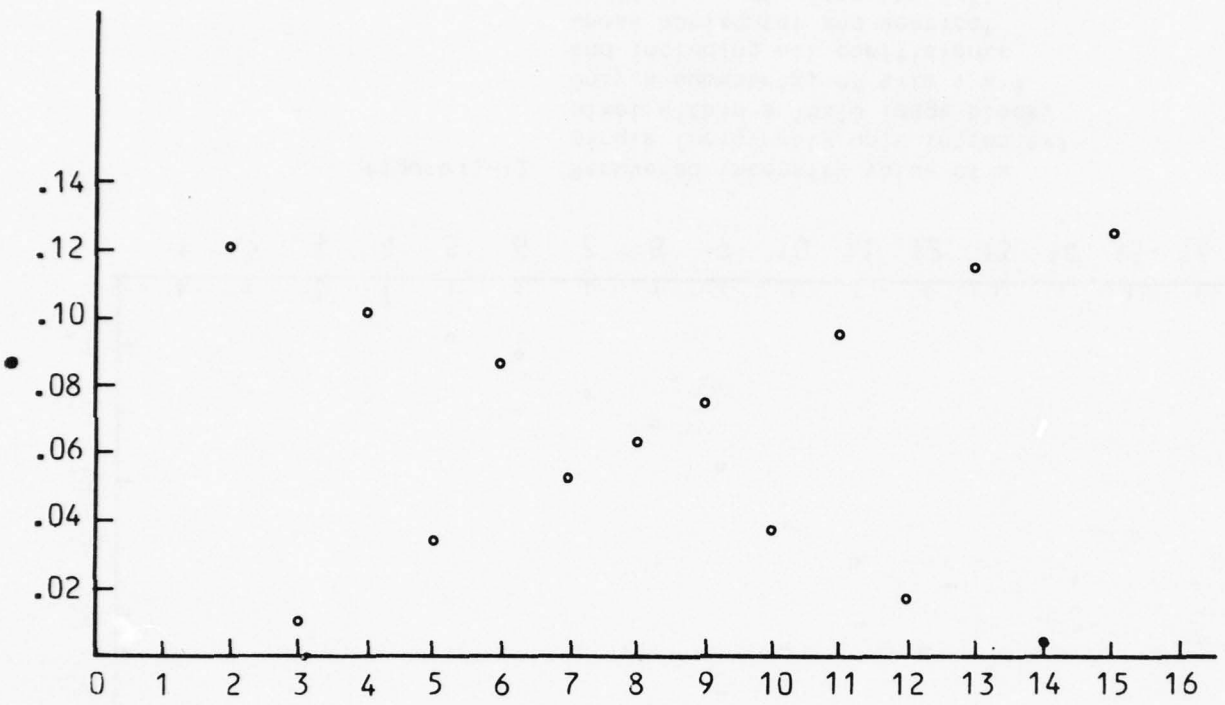


Figure II-11 Values of Cosine Transform coefficients, lying on the main diagonal of the transform array at locations (i,i), for a 16x16 image block containing a single non-zero-pixel of unit intensity.

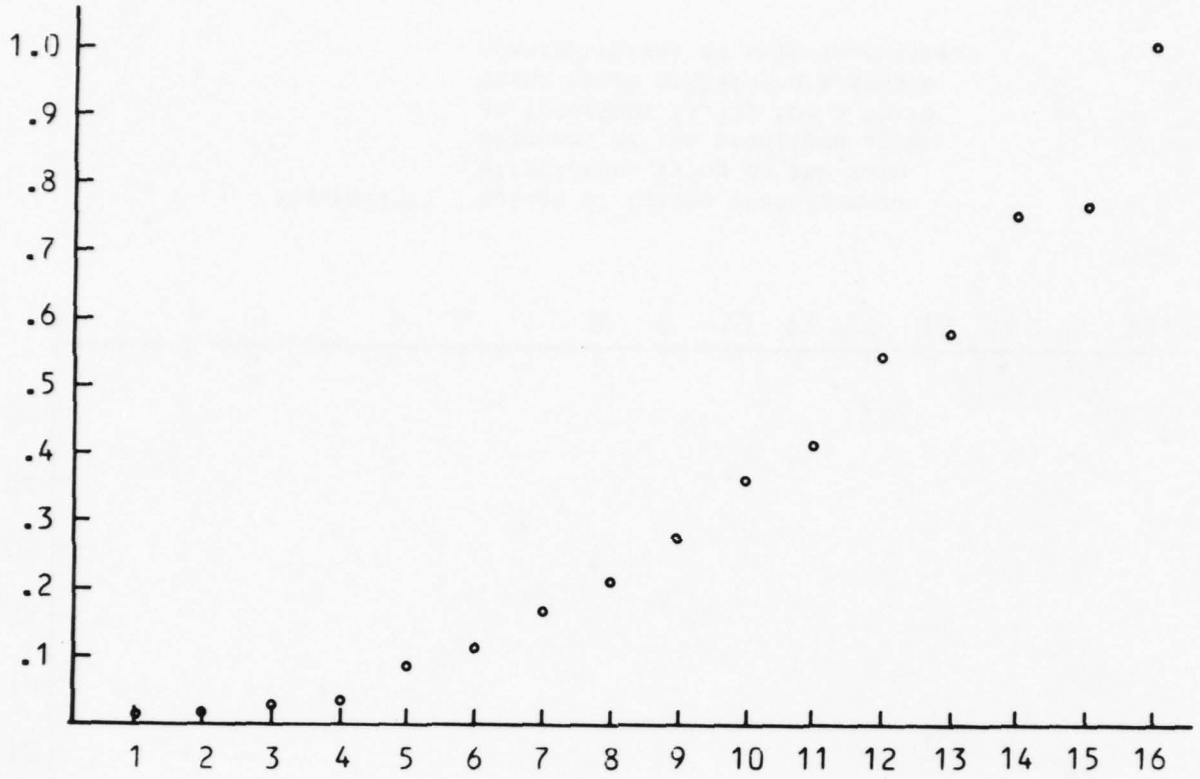


Figure II-12 Recovered intensity value of a single (originally unit intensity) pixel within a 16x16 image block. Only a submatrix, of size i x i and including all coefficients whose horizontal and vertical indices are <i, from the full 16x16 Cosine Transform matrix of the isolated pixel block is used for the inverse transformation.

Transform can be shown to produce a transform array of equal coefficient magnitudes, this is definitely not true of the Cosine Transform. While the transform energy is widely distributed, a few coefficients take on values several times those of others: particularly the highest spatial frequency at location (15,15) in the transform matrix. Fig. II-11 shows this and other higher frequency coefficients to be especially important. The Fourier transform could be expected to perform much better on this very special type of block.

In the spatial domain this very slow convergence of the Cosine Transform to a "boat in the water" is depicted by Figure II-12. Here inverse transforms are performed on submatrices of the full transform matrix of size 1x1, 2x2, 3x3, etc. until the full matrix is a used (16x16). Even at 15x15, when 88% of the coefficients are in use, the maximum absolute error exceeds 13%. Further experimentation with this threshold criterion shows that if the "boat" size is increased to 2x2 pixels, a 90% threshold detects these blocks with only a small loss in the overall compression rate.

While "city" blocks should clearly be provided additional coding bits in most adaptive schemes, blocks of a "coastline" or "boat" nature (described above) might best be represented through some other form of special handling, such as resorting to a spatial domain coding technique for these unique cases. Thus a means is still required to single them out. We have found the following criteria to be quite effective: all image blocks requiring a relatively small number of transform coefficients to account for, say, 90% of the total a.c. energy are found. Among these, several inverse transforms are performed on each transform array to determine those blocks needing a larger number of represented coefficients to achieve satisfactory pointwise convergence. (We use a maximum absolute error of 12.5 for

pixels in the range 0-255.) The first measure (a.c. energy) is basically a mean-square error one, while the second is a spatial domain maximum absolute error measure. Those blocks exhibiting good energy compaction (from a pure percentage viewpoint), while still showing poor spatial domain convergence are placed in the "boat-coastline" class. This has been found to work very well. "City" blocks can also be flagged reliably. These too exhibit slow pointwise convergence, but a large number of transform coefficients are needed to account for 90% of the total a.c. energy.

Adaptively increasing the number of coding bits allowed for a troublesome image block can lead to other problems in a photo reconnaissance context. The original of the "boat in the water" block used above appears at the left of Fig. II-13. After the coefficient quantization inherent in our zonal coding method is applied, and the result is inverse transformed, we obtain the block at the right of Fig. II-13. Quantization errors, as expected, have contributed to a heavy graininess in the water field surrounding the boat. Aesthetically this is not pleasing, but from an information-content point of view the boat "target" has been well represented. In fact, the quantization has caused its accentuation: The original boat intensity of 101 has been raised to 142. However potential false targets have been introduced. Partially surrounding the boat, at directions of 3, 9, and 12 o'clock, are three very dark pixels reminiscent of the "precursory undershoot" exhibited in the Fourier Gibbs phenomenon. These have been lowered in intensity from 72 to 44. They now deviate from the background intensity by an amount equal to the deviation of the original boat intensity. Aesthetically these dark pixels are simply a part of the overall objectionable graininess. From a reconnaissance viewpoint, however, their contrast as well as their symmetric placement about the boat could cause them

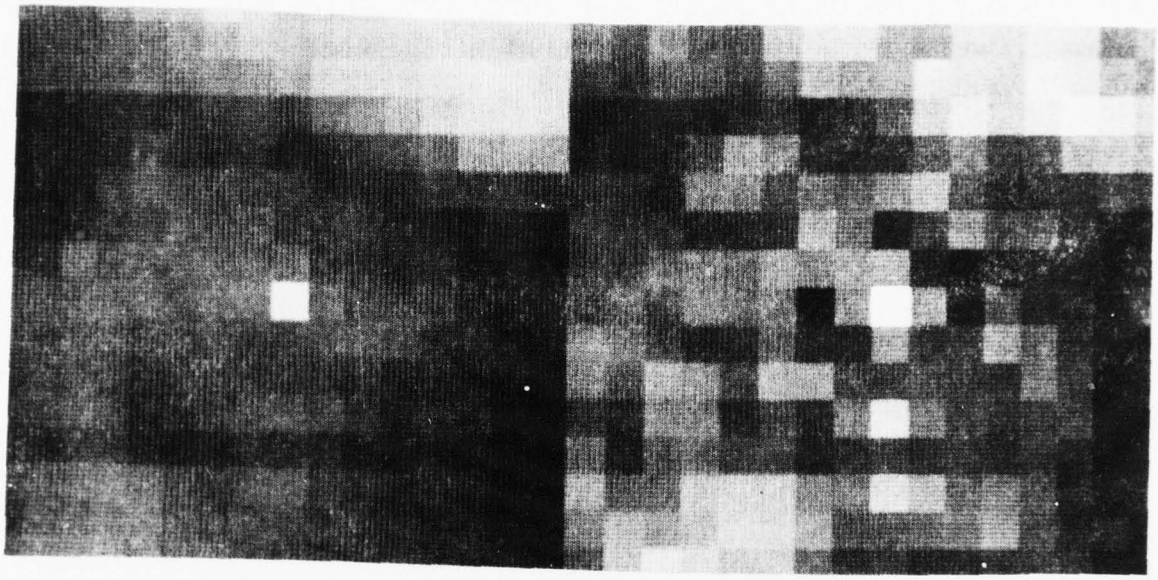


Fig. II-13 Quantization of frequency domain coefficients can result in graininess in the decoded image, but more importantly the introduction of false targets.

to appear as false targets: a problem almost as severe as the loss of a potential target discussed in connection with Fig. II-11.

A second problem arises when an adaptive coder assigns a significantly larger number of bits to a few blocks in an otherwise low frequency group of blocks. At the left of Fig. II-14, a water region, with two potential targets lying in distinct blocks, is shown. At the right we see the result of coding followed by decoding with a larger number of bits set aside for the two target blocks. The juxtaposition of two or more blocks of greatly differing graininess causes an aesthetic defect that disturbs even some professional photo analysts. The creation of artifacts that are potential false targets is certainly a legitimate issue. However when an entire water region appears grainy, the coding method causing this can actually be rated higher than another method that supresses grainularity except in blocks of high detail. Thus not all photo analysts are unconcerned with aesthetics, and this fact, possibly above all others, makes coding for photo reconnaissance specially difficult at higher compression rates.

A substantial portion of the graininess mentioned above is due to the large number of one bit (two level) quantizers prescribed for the higher frequency transform coefficients by the bit assignment algorithm. These coefficients have been provided no zero level and so any unnecessary coefficient, no matter how small, is forced to take on a full-scale (positive or negative) value equal to all others of the same bit-length. We have therefore, tested the efficacy of forcing all assigned two level coefficients to three levels. The result of this coding strategy change on "water" blocks is shown in Fig. II-15. Using strictly two-level quantizers on the higher frequency coefficients of the original (shown at the lower left) produces the graininess apparent in the decoded block at the upper left. Then in-

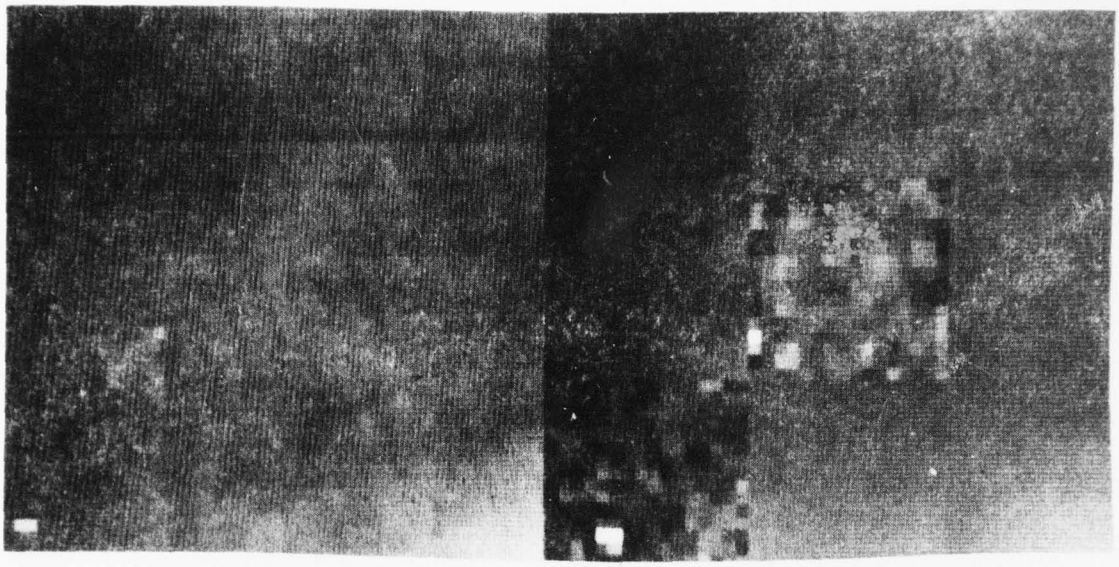


Fig. II-14 The justaposition of two inverse transformed blocks of greatly differing graininess is itself an aesthetic defect to which some photo analysis are sensitive.

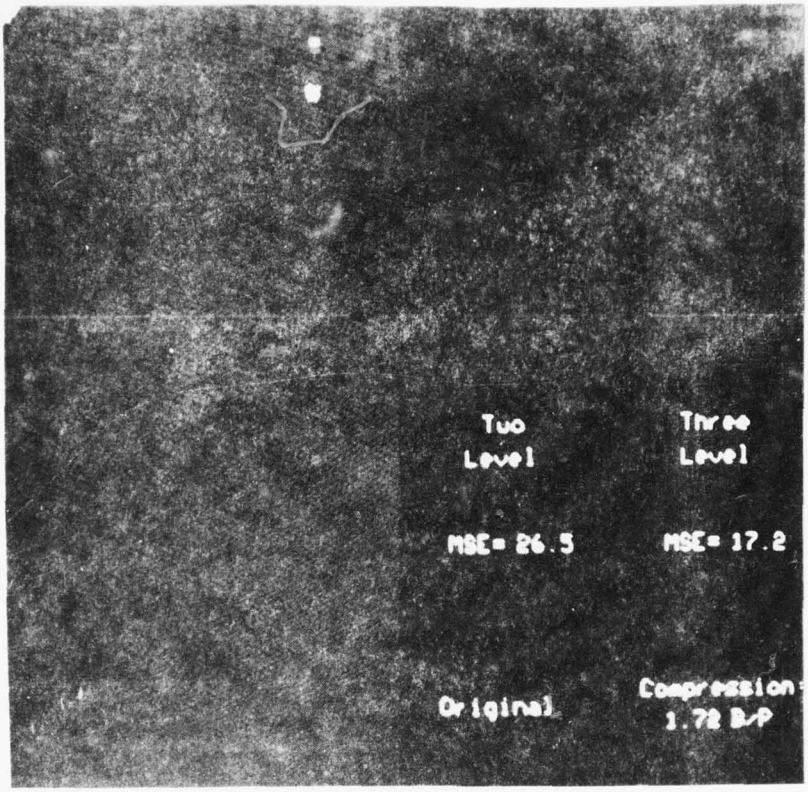


Fig. II-15 Use of three-level quantizers on blocks of low frequency information improves the graininess as well as m.s.e.

creasing all these two-level quantizers to three-level ones (and adjusting the bit assignment algorithm parameters to obtain the same overall compression), results in the much more satisfactory rendering of the upper right. Also, the mean-square error (m.s.e.) has dropped from 26.5 to 17.2.

Testing this same three-level strategy on blocks having three-pixel boats yields the results of Fig. II-16. Again the graininess is improved, and the m.s.e. is reduced from 39.3 to 24.3. With the large drop in graininess also comes a significant decrease in the probability of obtaining a false target.

If an isolated target block is similarly coded, however, Fig. II-17 shows that the target can almost completely disappear. At higher compressions, this disappearance is almost certain to occur. (In all these three-level tests we have used a compression rate of 1.72-1.77 bits/pixel). Somewhat surprisingly, though the graininess has again improved, the m.s.e. has increased from 66.3 to 84.7.

Use of the three-level strategy on "city" blocks renders these also more pleasing to the eye by reducing artifact clutter. The m.s.e. is not always improved, however. At 1.72 bits/pixel, the two-level coder applied to a block positioned over the right sections of two horizontal run-ways (see Fig. V-1), produces the block images of Fig. II-18. The m.s.e here has risen from 44.6 to 49.0. The same trend is in evidence at a higher compression rate of 0.59 bits/pixel (see Fig. II-19). The m.s.e. of 170.2 has become 198.3, and block boundaries have begun to appear.

We have concluded that three-level quantizers applied to higher frequency transform coefficients usually result in increased aesthetic appeal (at least at moderate compressions), but the possible loss of potential targets makes their use unjustifiable in reconnaissance work.

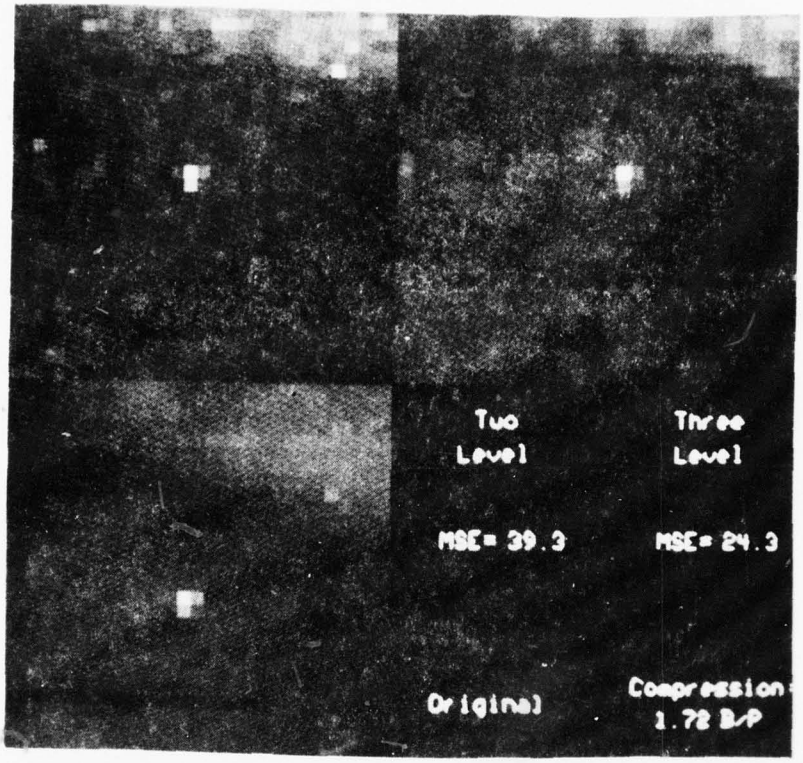


Fig. 11-16 Three-level quantizers are also beneficial when applied to blocks containing moderately small regions of detail

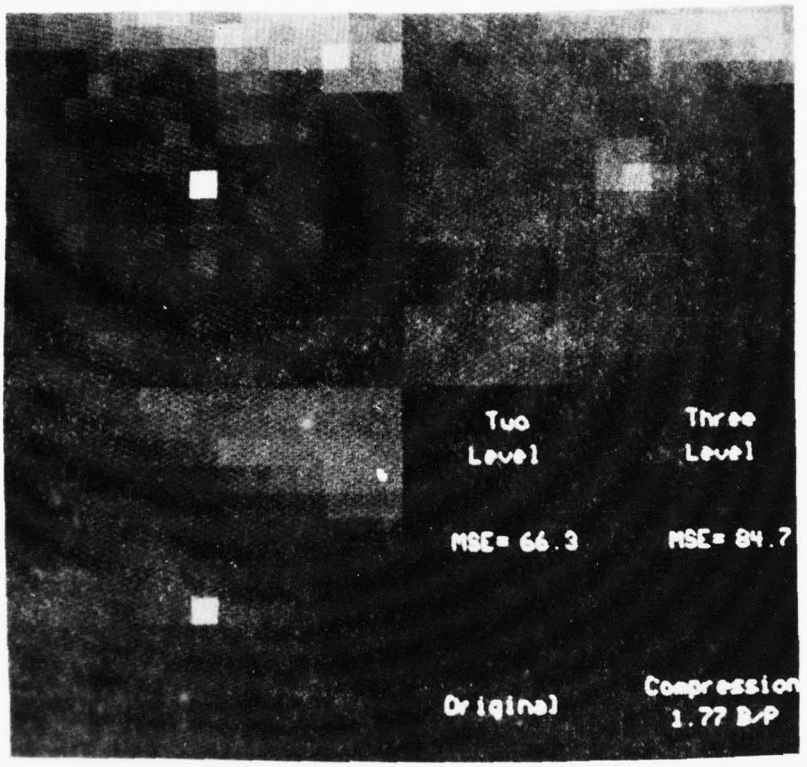


Fig. II-17 On isolated, single pixel target blocks, three-level quantizers can cause the obliteration of the target as well as a rise in the m.s.e.

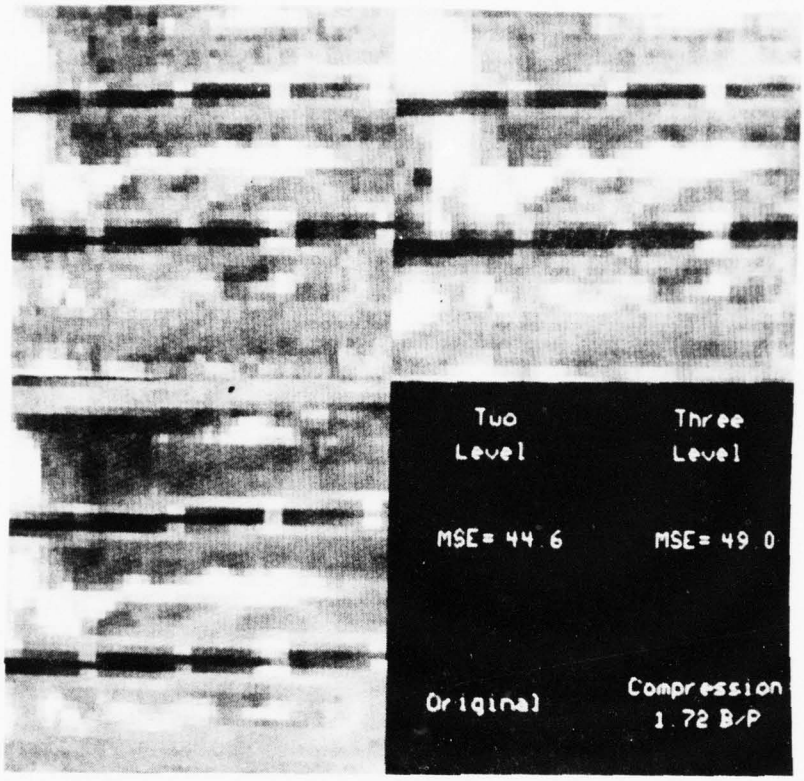


Fig. II-18 At 1.72 bits/pixel, the application of three-level quantization to "city" blocks typically reduces artifact clutter. The m.s.e. can increase, however.

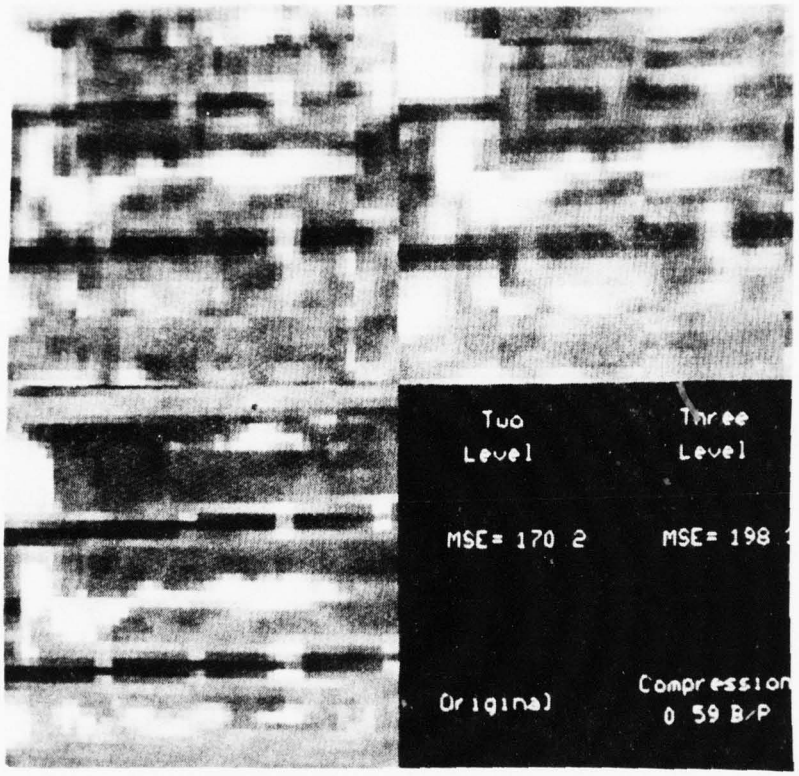


Fig. II-19 At 0.59 bits/pixel, the three-level quantization scheme may not far as well. The m.s.e. again increases, and block boundaries are accentuated.

Transmission errors in transform coded information is conventionally thought to be preferable to the "salt and pepper" errors accompanying the use of some spatial domain coders such as PCM. We do not believe this issue is at all clear cut in the case of aerial reconnaissance photography. As for frequency domain information, the toggling of a code bit may or may not cause large decoding errors. But the results of channel errors applied to some spatial domain coders will often be highly localized ("salt and pepper"), and so frequently the resulting artifacts can be ruled out as targets by contextual considerations. In the last analysis, these aesthetically annoying defects would then be of no importance. But toggling bits of higher significance in transform coefficient information can soften target edges throughout a block thus impeding target identification and even detection. Since each coefficient affects every pixel value within an inverse transformed block, an extreme error situation can result in the obliteration of several targets within a block. This latter is not common, however.

III. Hybrid Coding

This project was designed to investigate the use of a hybrid method in the transform coding of aerial reconnaissance images.

The hybrid technique that was investigated was introduced by Habibi [7]. It consists of taking a one dimensional unitary transform of a strip of data and passing it into a bank of differential pulse code modulators, DPCM. The resulting differential signal is then quantized and passed through a simulation of a noisy channel. At the receiver the data is then fed into an associated DPCM system, and the inverse transform is taken.

There were four main areas that were investigated concerning the implementation of the method:

1. Transform Method

2. Optimum Strip Length

3. Quantization Method

4. DPCM System Configuration

III.1. Transform Method

There are many transform techniques in use which are available for this method. Among these are: Karhunun-Loeve, (KL), Fourier, Walsh, and cosine. These are all discussed in Ahmed and Rao [8]. The KL transform is recognized as being the best in the MSE sense when fewer than the full set of basis functions are used, as is necessary in image compression. However it is rarely used, as no fast algorithm for its computation has been developed.

The cosine transform was introduced by Ahmed, Natarajan, and Rao [3]. They show that its performance is a close approximation of the KL transform. The hybrid method implemented here uses the discrete cosine transform. Chen, Smith, and Fralick [4], have published a fast algorithm for computing the cosine transform that does not require the use of a fast Fourier transform. This algorithm was used to further reduce the computational cost of the hybrid method.

III.2. Strip Length

When dealing with large images it would be ideal to be able to work with the whole image at one time. This would take advantage of all of the redundancy available in the image. However the amount of computational time and computer storage space that this would require would be prohibitive. Therefore the picture must be broken up into smaller subunits which are processed separately. In implementing this method a decision had to be made as to the desired length of the strips over which the one dimensional transforms would be made. Figure III-1 presents the relationship between the strip length and the operations count required to compute the fast DCT.

This is a significant portion of the total number of computation required for the method and it is desired to keep this as low as possible in order that this method might be realizable in a "real time" system with a dedicated processor.

The method was implemented using strips of length ranging from four to 32 pixels long. The resultant MSE performance as a function of the strip size used is graphed in figures III-2 and III-3. As can be seen there is no appreciable increase in MSE performance for strips with length greater that eight pixels, at 1.5 bits/pixels. However, at 0.5 bits/pixels, the MSE of images processed with strips eight pixels long is significantly larger than those with strips sixteen pixels long. There is also significant improvement in the visual characteristics of images processed with strips sixteen pixels long, versus those processed with strip lengths of eight pixels. No similar improvement, when compared with the increased computational cost, is noticed when the strip size is increased to 32 pixels.

III.3. Quantization Methods

The images that were processed were originally quantized to 256 levels, or eight bits/pixel. In order to achieve the desired bandwidth reduction the transmitted image was required to be sent at an average rate which varied between 1.6 and 0.5 bits/pixel, depending on the degree of compression desired.

The optimal bit assignments for the method depends on the type of image being processed. Images which change rapidly in the spatial domain contain a large amount of high frequency information. It is therefore desirable to assign a relatively larger number of bits to the higher coefficients for such an image, than an image which changes slowly in the spatial domain. In order to determine the optimum bit assignments it is first

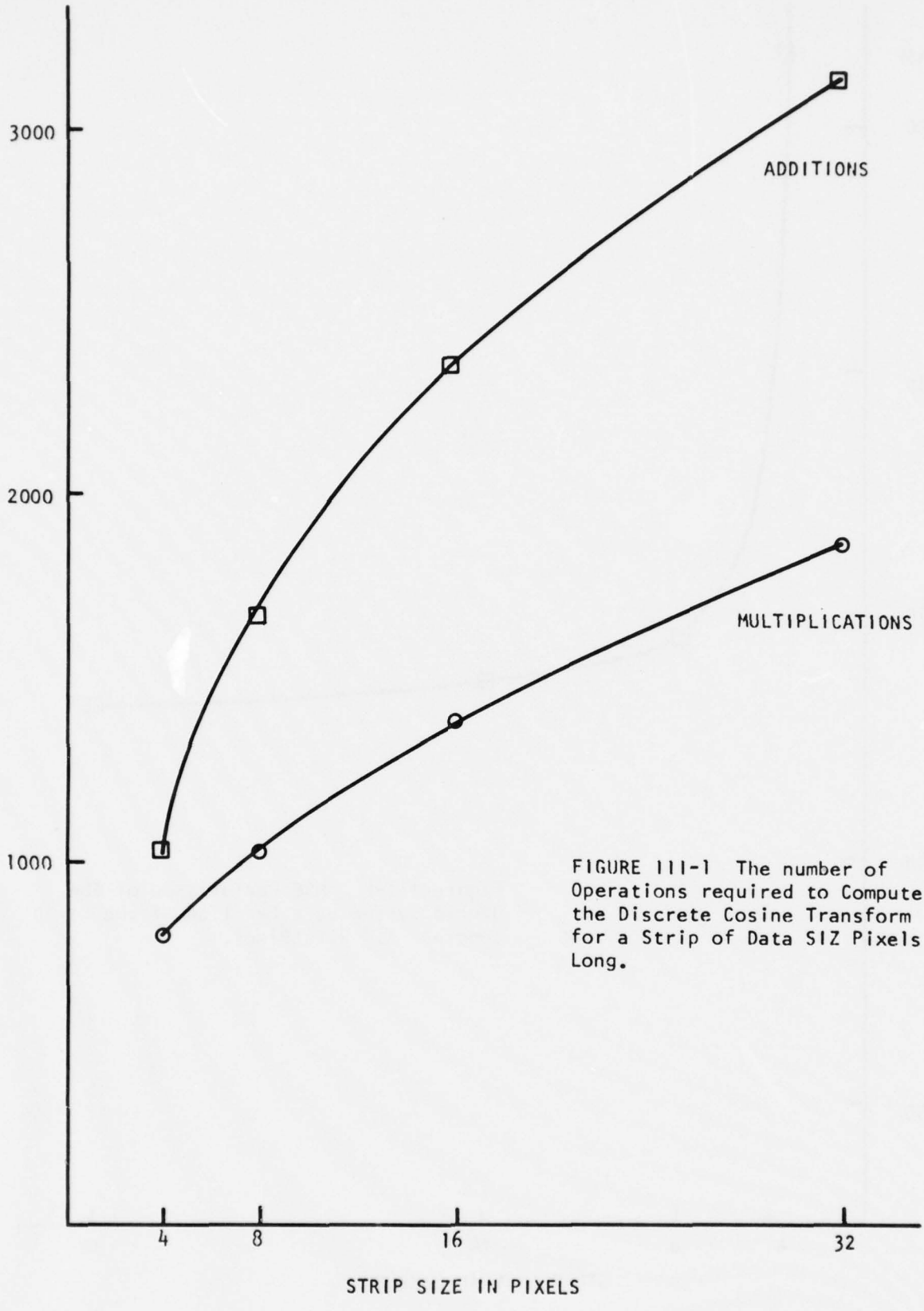


FIGURE III-1 The number of Operations required to Compute the Discrete Cosine Transform for a Strip of Data SIZ Pixels Long.

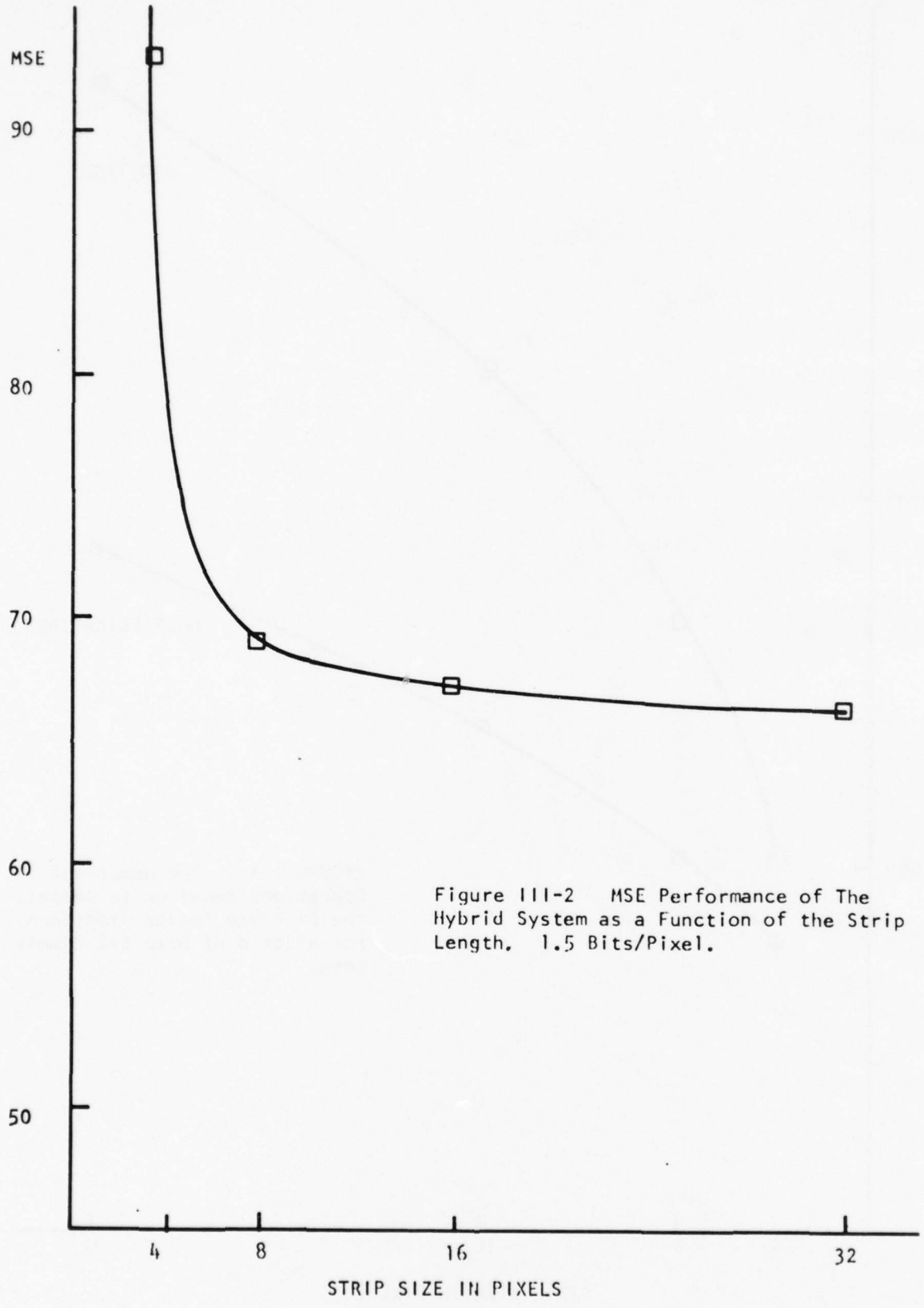


Figure III-2 MSE Performance of The Hybrid System as a Function of the Strip Length. 1.5 Bits/Pixel.

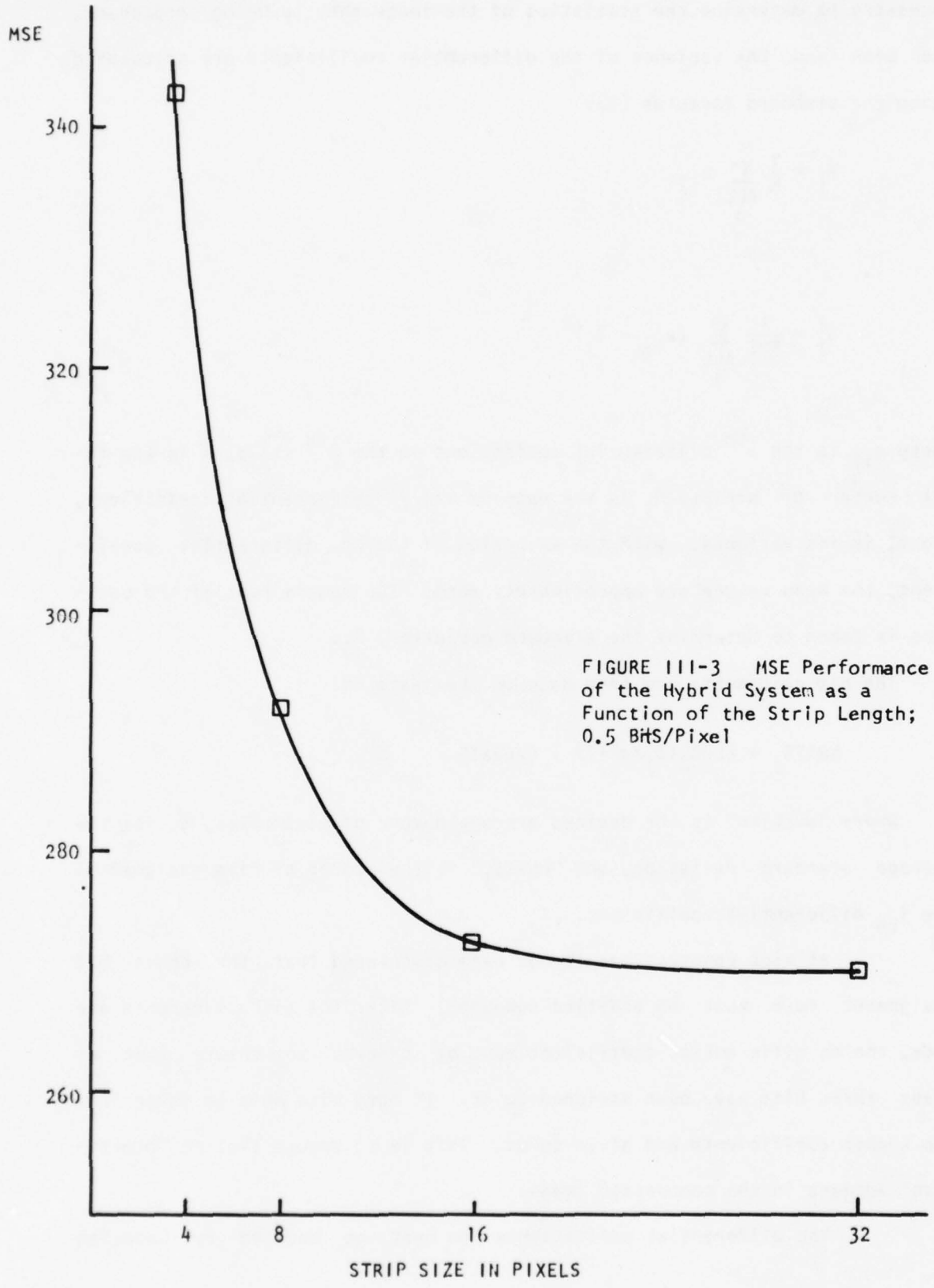


FIGURE III-3 MSE Performance of the Hybrid System as a Function of the Strip Length; 0.5 BitS/Pixel

necessary to determine the statistics of the image that is being processed. The mean and the variance of the differential coefficients are calculated using the standard formulas [9]:

$$\hat{\eta}_j = \frac{1}{M} \sum_{\substack{ALL \\ i}} a_{ij}$$

$$\hat{\sigma}_j^2 = \frac{1}{M-1} \sum_{\substack{ALL \\ i}} (a_{ij} - \hat{\eta}_j)^2$$

Where a_{ij} is the j^{th} differential coefficient in the i^{th} strip, m is the total number of strips, $\hat{n}_j$ is the mean of the j^{th} differential coefficient, and $\hat{\sigma}_j^2$ is its variance. With the exception of the dc differential coefficient, the mean values are approximately zero. The square root of the variance is taken to determine the standard deviation, $\hat{\sigma}_j$.

The bit asignments are then made on the basis of:

$$NBITS_j = [LOG_2(\hat{\sigma}_j/\overline{\sigma})+1] \times AVEBITS$$

Where "avebits" is the desired average number of bits/pixel, $\overline{\sigma}$ is the average standard deviation, and "$nbits_j$" is the number of bits assigned to the j_{th} differential coefficient.

At high compressions it has been discovered that the above bit assignment rule must be modified somewhat. After the bit assignments are made, the dc differential coefficient must be checked to ensure that at least three bits have been assigned to it. If not, bits must be taken from the higher coefficients and given to dc. This is to ensure that no "blockiness" appears in the compressed image.

The differential coefficients can best be modeled by Lapacian

random variables (with a two sided exponential density function). As described by Max [10], the noise introduced by the quantizer may be minimized in the MSE sense by using non-uniform quantization steps. These cutoff levels were determined using the iterative technique that Max described.

The first line must be sent via a pulse code modulation technique. It was assigned eight bits per coefficient, due to the greater amount of information which it contains.

III.4. DPCM System Configuration

One of the design constraints for the system was that it was to perform well in the presence of simulated channel noise. There are two possible methods for the error protection of the DPCM system: updating and leak.

In the updating method the original signal, rather than the differential one, is periodically transmitted at a much higher bit rate. Thus the quantizer is periodically restarted, which would prevent a channel error from propagating through the entire image. This method was implemented by sending the original spatial information every 32nd line, quantized to six bits/pixel. In the resulting image the channel errors appeared as black and white streaks running down a column of the image. At low bit rates, 0.5 bits/pixel, the updating lines were much sharper, and thus distracting. Also at low bit rates the cost of such updating, as measured in the decrease in the available bits for the differential lines, becomes prohibitive. Some improvement was noticed when the cosine transform coefficients, rather than the grey levels, of the updating lines were transmitted. This served to smooth the error out over a larger region of the image, but the cost is still prohibitive.

The leak method can be designed in terms of the optimal MSE predictor. It is desired to determine the best MSE estimate of the signal, s_{ij},

in terms of the previous value of the signal, $s_{i-1,j}$.

The signal to be transmitted will then be the difference between this estimate and the actual signal. The estimate, $\hat{s}_{ij}$, of s_{ij} is:

$$\hat{s}ij = \alpha_j s_{i-1,j}$$

We desire $E\{(s_{ij}-\hat{s}_{ij})^2\}$ to be minimum. Therefore we wish to minimize:

$$E\{(s_{ij}-\alpha_j s_{i-1,j})(s_{ij}-\alpha_j s_{i-1,j})\}$$

Differentiating with respect to α_j and equating to zero:

$$-2E\{s_{ij}s_{i-1,j}\}+2\alpha_j E\{s_{i-1,j_2}\} = 0$$

$$\alpha_j E\{s_{i-1,j_2}\} = E\{s_{ij}s_{i-1,j}\}$$

$$\alpha_j = E\{s_{ij}s_{i-1,j}\}/E\{s_{i-1,j_2}\}$$

Using the following estimates for the autocorrelation, $R_j(0,0)$, of the j^{th} cosine transform coefficients, and the crosscorrelation, $R_j(0,1)$, between that coefficient and the corresponding one in the previous line:

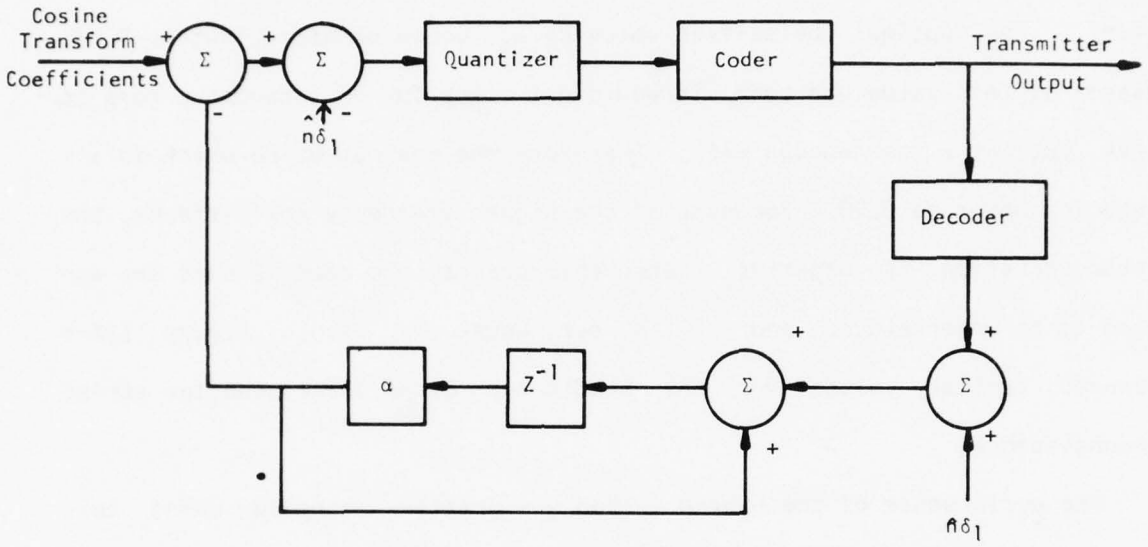
$$R_j(0,0) = \frac{1}{M}\sum_{\substack{ALL\\ i}} s_{ij}^2$$

$$R(0,1) = \frac{1}{M}\sum_{\substack{ALL\\ i}} s_{ij}\ s_{i-1,j}$$

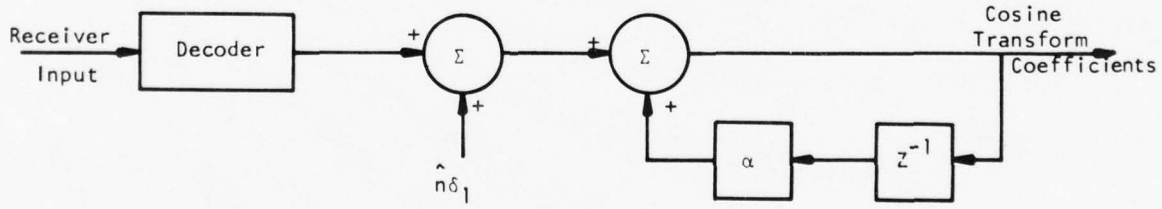
$$\alpha_j = R_j(0,1)/R_j(0,0)$$

Figure III-4 presents a typical transmitter-receiver pair for this method. This method would require a parallel set of these systems, equal in number to the strip length. Notice that the mean value of the differential information must be taken into consideration for the dc coefficients. This is due to the high degree of correlation which is present between the dc values. The optimal noise-free value of α_1 would be approximately 0.98. However if this value was used, it would not allow for any channel errors to "leak out" at a fast enough rate. Therefore the maximum value which is allowed for any α is 0.90. For some of the higher frequency coefficients, the crosscorrelation is negative. When this occurs, the coefficients are assumed to be uncorrelated, and α is set equal to zero. Figure III-5 presents typical values for the statistics of an image used for aerial reconnaissance.

The performance of the hybrid method was greatly improved using this DPCM system. The resultant images were much sharper, and the method was less sensitive to channel noise. Figure III-6 shows the performance of this method in the presence of channel noise. Images suitable for reconnaissance analysis are obtainable with noise probability up to 10^{-3}, at 1.5 bits/pixel.



(a)



(b)

FIGURE III-4 Hybrid Transmitter and Receiver

PICSIZ= 256
BLKSIZ= 16
CR= 1.43870
DC MEAN= 181.27109

I	STDEV	BITS	AUTCOR	CRSCOR	ALPHA
1	144.73854	4	3685104.25000	3664562.50000	0.90000
2	96.65750	3	41299.50000	36244.19141	0.87759
3	81.53075	3	19969.64648	16295.15723	0.81518
4	64.10664	3	10182.71875	7869.81494	0.77286
5	58.29052	2	6322.01855	4277.67041	0.67663
6	49.87075	2	4642.27930	3154.22412	0.67946
7	44.30973	2	3273.26440	2064.99561	0.63087
8	38.83828	1	2301.63501	1353.48926	0.58806
9	33.83505	1	1775.19189	1047.20959	0.58991
10	28.62680	1	1071.29919	515.02222	0.48075
11	26.96406	1	1052.92200	587.34009	0.55782
12	23.55284	0	734.08966	360.67136	0.49132
13	21.31132	0	545.34979	222.73289	0.40842
14	19.40464	0	457.43243	191.13969	0.41785
15	18.80035	0	413.49615	157.10326	0.37994
16	17.45918	0	327.34961	86.74522	0.26499

Figure III-5

HYBRID STATISTICS FOR A TYPICAL
RECONNAISSANCE IMAGE

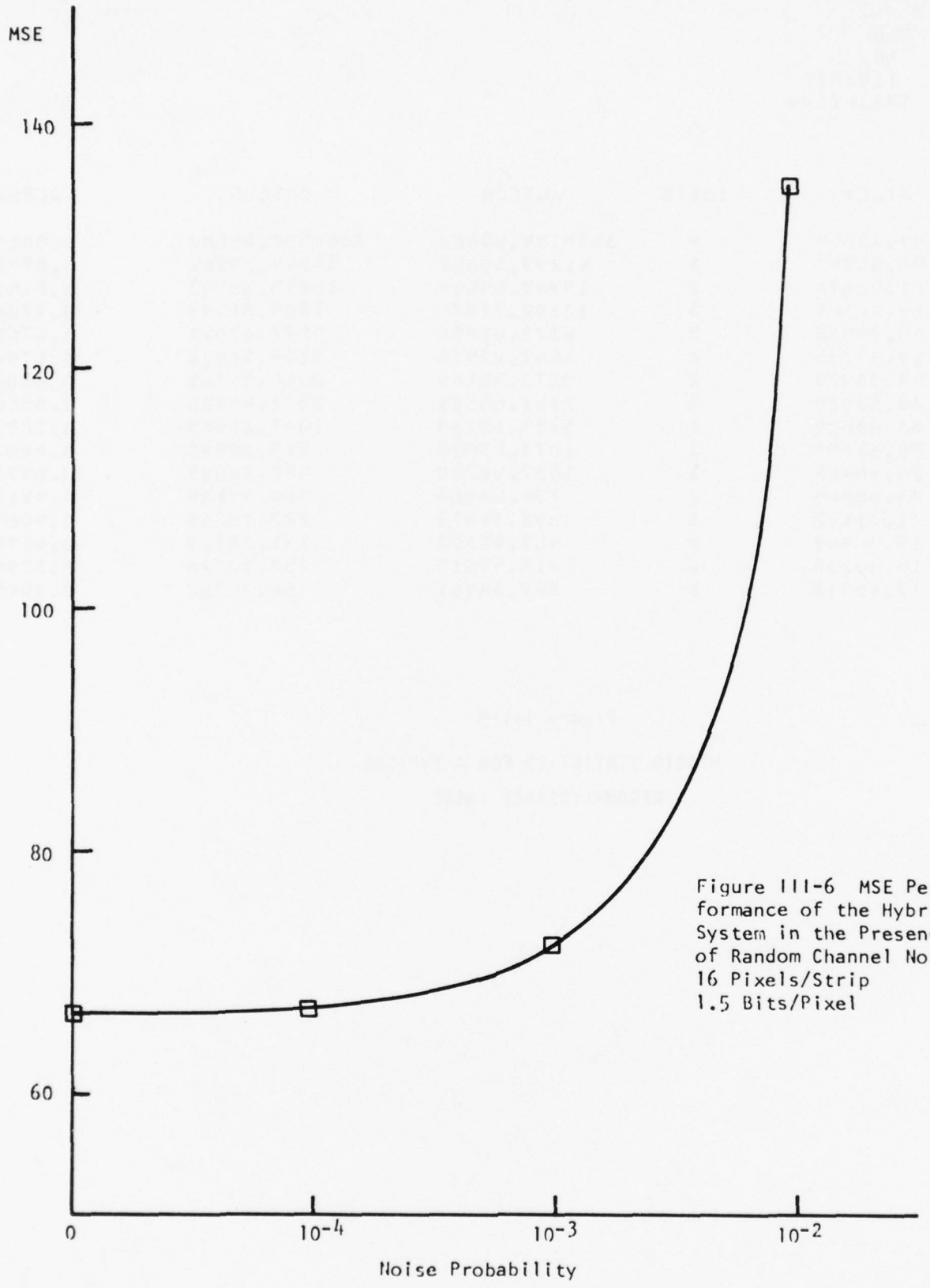


Figure III-6 MSE Performance of the Hybrid System in the Presence of Random Channel Noise.
16 Pixels/Strip
1.5 Bits/Pixel

III.5. Summary

This project was designed to investigate and implement a hybrid method for image transform coding. The results show that a hybrid method, using a one dimensional discrete cosine transform, and DPCM between adjacent lines, can produce results that are comparable to other available techniques. The advantages of the method are its speed of operation, and its relative ease of implementation. The hybrid system produces images of the quality needed for reconnaissance photo analysis at compressions down to 1.5 bits/pixel. The system also performed extremely well in the presence of random channel noise. This method could also be used in other applications, such as digital television, were a rapid method of image compression is needed.

IV. Block Truncation Coding

The use of Block Truncation Coding (BTC) relative to this work was first presented in an earlier report [1] and will be reviewed briefly here. Block Truncation Coding can be formulated as the application of a nonparametric (one-bit) moment preserving quantization. The BTC algorithm originally presented preserved only the first two sample moments in each 4 x 4 image block. The threshold was chosen a priori as the sample mean. In this section, we will present our latest modifications of BTC. These modifications along with a thorough description of the basic algorithm can be found in [11]. A theoretical development is presented in [12]. Before presenting the modifications to BTC we will present two other nonparametric one bit quantization schemes using classical fidelity criteria that we used to compare with BTC along with a brief review of BTC. Let $m=n^2$ and let $X_1, X_2, \ldots X_m$ be the values of the pixels in a block of the original picture. (n=4 for our case)

Let

$$\overline{X} = \frac{1}{m} \sum_{i=1}^{m} X_i \text{ be the first sample moment}$$

$$\overline{X^2} = \frac{1}{m} \sum_{i=1}^{m} X_i^2 \text{ be the second sample moment} \qquad \text{(IV-1)}$$

$$\overline{\sigma^2} = \overline{X^2} - \overline{X}^2 \text{ be the sample variance}$$

As with the design of any one bit quantizer (see Figure IV-1), it is necessary to find a threshold and two output levels for the quantizer such that:

$$\text{if } X_i \geq X_{th} \qquad \text{output} = b \qquad \text{(IV-2)}$$

$$\text{if } X_i < X_{th} \qquad \text{output} = a$$

$$\text{for } i = 1,\ldots,m.$$

where

X_{th} is the threshold

a and b are the "low" and "high" output levels respectively.

For our basic BTC quantizer we shall make an ad hoc assumption that $X_{th} = \overline{X}$. This seems reasonable; however, we will later modify this assumption to get a more consistent result. The output levels a and b for a two-level moment preserving quantizer are found by solving the following equations:

let q = number of X_i's greater than X_{th} ($\overline{X}$ in this case)

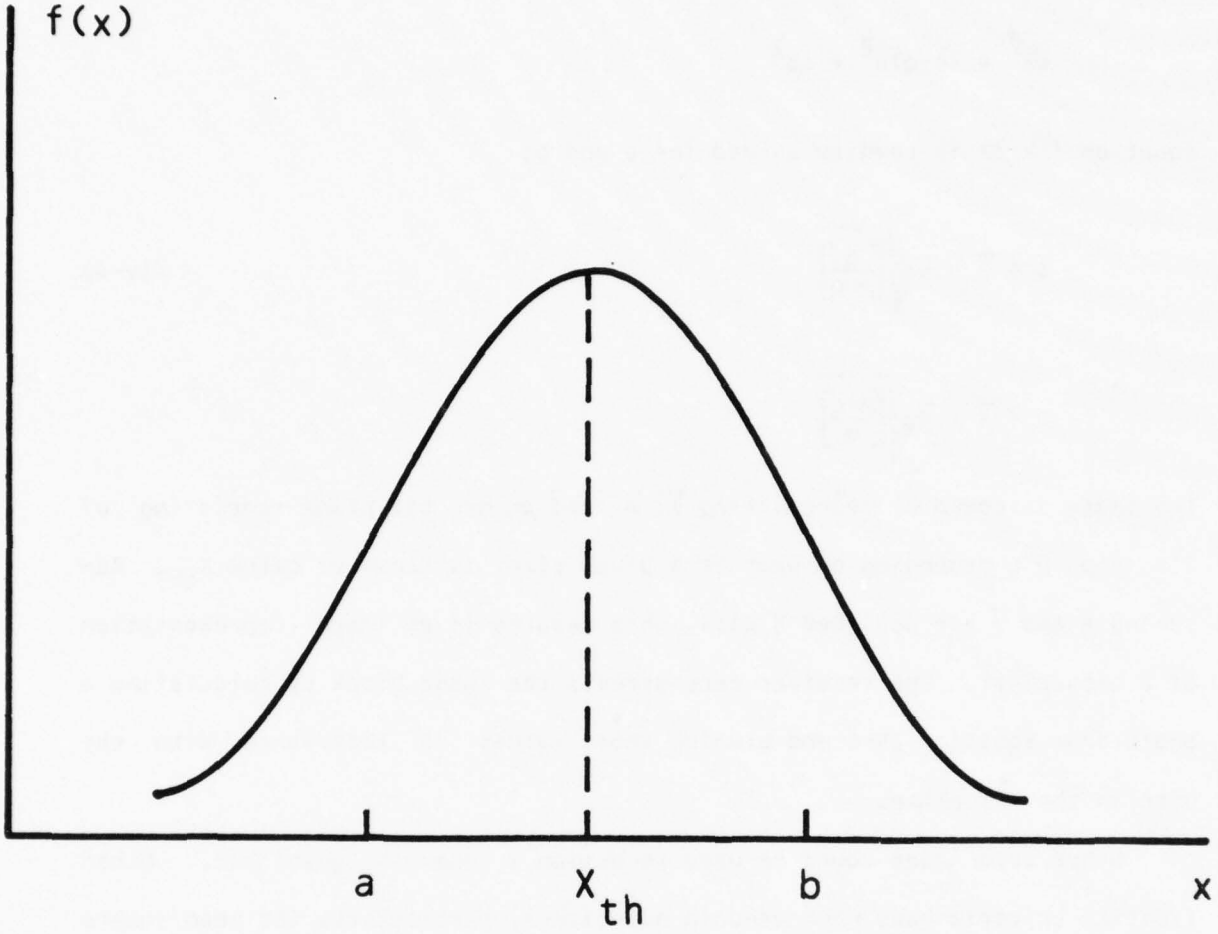


Figure IV-1 In designing a one bit quantizer given the data values (assumed to be a continuous density function), one must find a threshold value X_{th} and two output levels a and b, respectively.

We then have

$$m\overline{X} = (m-q)a + qb \qquad \text{(IV-3)}$$

$$m\overline{X^2} = (m-q)a^2 + qb^2$$

Equation (IV-3) is readily solved for a and b:

$$a = \overline{X} - \overline{\sigma}\sqrt{\left[\frac{q}{m-q}\right]} \qquad \text{(IV-4)}$$

$$b = \overline{X} + \overline{\sigma}\sqrt{\left[\frac{m-q}{q}\right]}$$

The image is coded by transmitting $\overline{X}$, $\overline{\sigma}$ and an nxn bit plane consisting of 1's and 0's depending on whether a given pixel is above or below X_{th}. Assuming $\overline{X}$ and $\overline{\sigma}$ are assigned 8 bits, this results in an image representation of 2 bits/pixel. The receiver reconstructs the image block by calculating a and b from Equation IV-4 and placing those values in accordance with the bits in the bit plane.

Other techniques could be used to design a one-bit quantizer. Other fidelity criteria have been used in quantizers, particularly the mean square error (MSE) and mean absolute error (MAE). BTC uses a fidelity criterion of moment preservation (MP). To use the MSE fidelity criterion, one proceeds by first constructing a histogram of the X_i's (i.e., sorting the X_i's). Let Y_1, Y_2, ... Y_m be the sorted X_i's; i.e., $Y_1 \leq Y_2 \ldots \leq Y_m$. Again let q be the number of X_i's greater than X_{th}. Then a and b are found by minimizing:

$$J_{MSE} = \sum_{i=1}^{m-q-1} (Y_i - a)^2 + \sum_{i=m-q}^{m} (Y_i - b)^2 \qquad \text{(IV-5)}$$

where

$$a = \frac{1}{m-q} \sum_{i=1}^{m-q-1} Y_i$$

$$b = \frac{1}{q} \sum_{i=m-q}^{m} Y_i$$

In general it is impossible to solve this equation in closed form for X_{th}, a, and b. One way to solve this problem is to try every possible threshold (there are at most m-1 thresholds) and pick the one with smallest J_{MSE}. Assuming a and b have 8-bit resolution, this gives a representation of 2 bits/pixel.

The problem of using the MAE fidelity criterion is very similar to the MSE. The values a and b are found by minimizing:

$$J_{MAE} = \sum_{i=1}^{m-q-1} \left|Y_i - a\right| + \sum_{i=m-q}^{m} \left|Y_i - b\right| \qquad \text{(IV-6)}$$

where

$$a = \text{median of } (Y_1, Y_2, \ldots Y_{m-q-1})$$

$$b = \text{median of } (Y_{m-q}, \ldots Y_m)$$

Here again the quantizer is arrived at by an exhaustive search. Results using these quantizers and BTC are shown in Figure IV-2. The performance of BTC is quite good when compared to these standard fidelity criteria. The

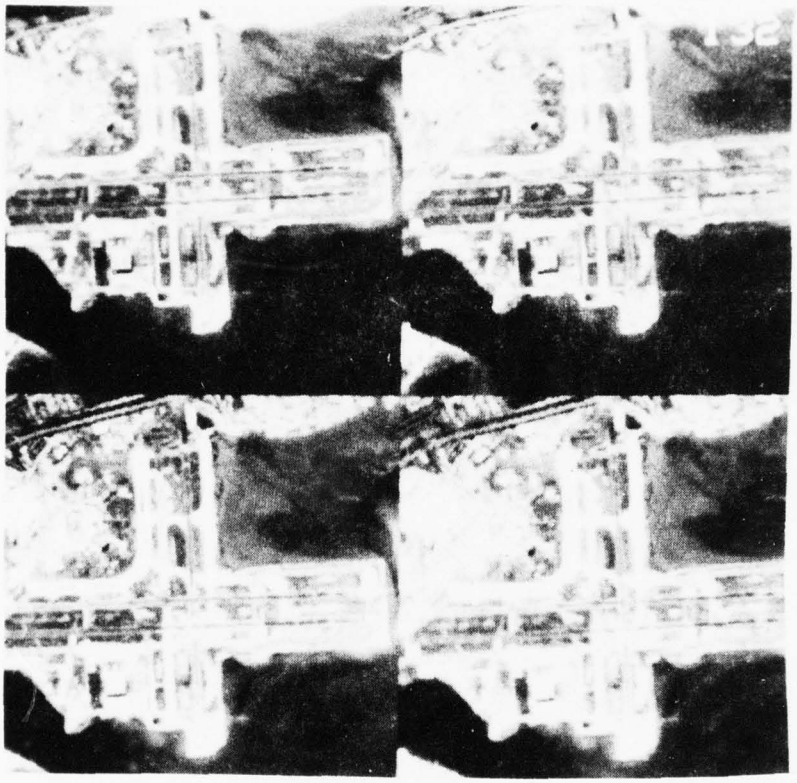


Fig. IV-2: Results using various fidelity criteria. All representations are 2.0 bits/pixel. Upper left: MSE; Upper right: MAE; Lower left: MP; Lower right: MSE and also assuming image data uniformly distributed with each block.

advantage of just preserving the sample moments is obvious because an exhaustive search is not necessary to match the criterion.

In some cases it is possible to make an ad hoc assumption of the probability density of the image in each block. Once the density function is known (or guessed), the quantizer is immediately specified using either MP, MSE, or MAE fidelity criteria. The results obtained using this procedure are usually quite poor when high resolution imagery is used. This procedure can sometimes be used quite successfully in television quality "head and shoulders" pictures. The results are poorer and more coding artifacts can be seen in the reconstructed imagery.

One of the disadvantages of BTC is that the compression achieved corresponds to only 2 bits/pixel. In many image coding schemes it is desired to obtain data rates in the range of 1.0-1.5 bits/pixel.

As mentioned above, it is necessary to transmit some overhead information for the quantizer in each block. The information usually transmitted is $\overline{X}$ and $\overline{\sigma}$. One obvious way of lowering the number of bits for image representation is to assign less than 8 bits to $\overline{X}$ and $\overline{\sigma}$. Experimental evidence has indicated that it is possible to code $\overline{X}$ with 6 bits and $\overline{\sigma}$ with 4 bits. This allows for considerable savings and few perceivable errors upon reconstruction. This then gives a representation of 1.63 bits/pixel. Alternately a and b (instead of $\overline{X}$ and $\overline{\sigma}$) could be transmitted to the receiver and assigned fewer bits. Experimental evidence indicates that the representation obtained at the receiver is better if $\overline{X}$ and $\overline{\sigma}$ are transmitted and a and b computed at the receiver; i.e., the mean needs more precision than the contrast (standard deviation) for accurate perception.

By choosing the threshold of the quantizer at $\overline{X}$, it has been observed that partitioning of the data leads to some "unnatural" appearance of the

data. For high resolution imagery, this manifested itself by some unacceptable coding artifacts. It would be better if somehow the fidelity criterion allowed for automatic threshold selection as does MSE and MAE. This can be arrived at by forcing the quantizer to preserve not only the first two sample moments but also higher moments. A detailed development is presented in [11]. This method requires extra computation at the transmitter, but the receiver is not affected. It should be mentioned that this method of automatic threshold is far easier than the MSE or MAE quantizers discussed above since an exhaustive search is not necessary to find a, b, and q.

This new threshold technique improved the subtle features (such as near edges) of the image that are usually important in analysis of aerial photography imagery. For some of the imagery used in this study the coding artifacts produced using the sample mean as the threshold were such that the photo interpreters rated the images poorer. When the third moment preserving technique was used many of these coding artifacts disappeared although the mean square error was not significantly changed.

As with all non-information preserving image coding, coding artifacts are produced in the image. It became apparent very early in this study that BTC produces artifacts that are very different from transform coding. These artifacts are usually produced in regions around edges and in low contrast areas indicated by a sloping gray level. As mentioned above, BTC does produce sharp edges. However, these edges do have a tendency to be ragged. Transform coding usually produces edges that are blurred and smooth. The second problem in low contrast regions is due to inherent quantization noise in the one bit quantizer. Here sloping gray levels can turn into false edges. It should be emphasized that these coding artifacts are problems in high resolution aerial reconnaissance images where man-made objects are im-

portant (i.e., edges). These coding artifacts usually are not any problem in television quality "head and shoulders" imagery.

One of the problems that BTC has is that it is really a one-dimensional quantization technique. In no way does BTC exploit the two-dimensional nature of the image within each block as does most other forms of image coding. Also BTC generally has a poor response near the spatial frequency of 1/2 cycle per block.

One method to improve both of the problems above is a hybrid formulation. First a highly compressed Cosine transform coded image is subtracted from the original image. For the results presented here the transform picture was obtained by taking the two-dimensional Cosine transform over 16 x 16 pixel blocks. Only the eight non d.c. coefficients in the upper left section of each block were retained. This corresponds to a zonal filtering method. This lead to a representation of 0.25 bits/pixel for the highly compressed image. BTC is then used on this difference picture and the recombination formed at the receiver. While this does increase the computational load, the improvement seems to be significant enough to give this method further attention. Figure IV-3 presents results of this hybrid method. This technique exploits the edge preservation of BTC and helps in the low contrast regions of the image by improving the frequency response. Recently Texas Instruments has done a study of implementing BTC on an integrated circuit chip using VLSI techniques. This study indicates BTC could be implemented with a gate count of 3800 and a maximum delay path of 30 gates [16].

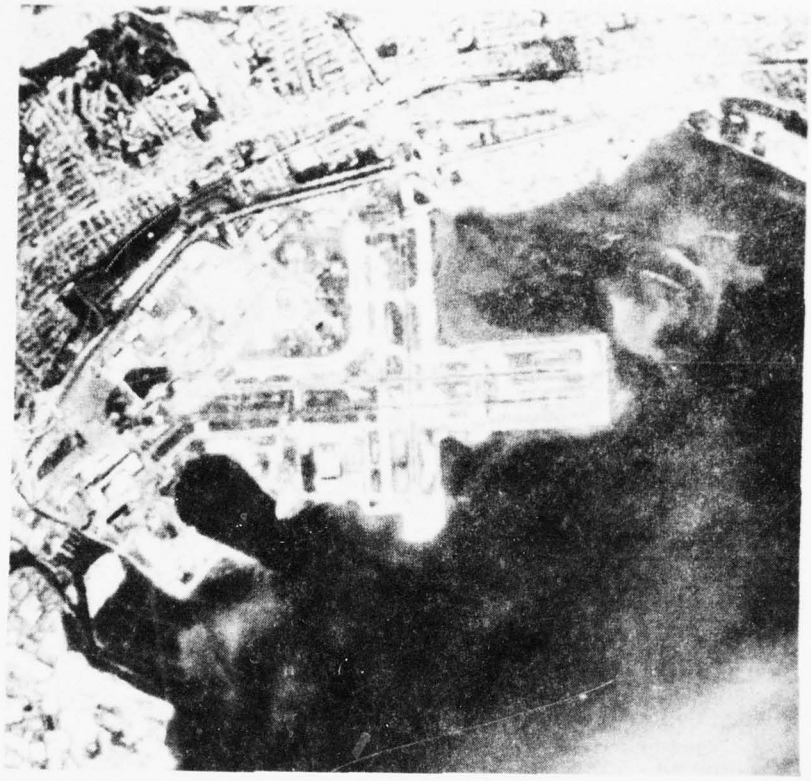


Fig. IV-3: Results of Hybrid Formulation of BTC. Representation is 1.88 bits/pixel.

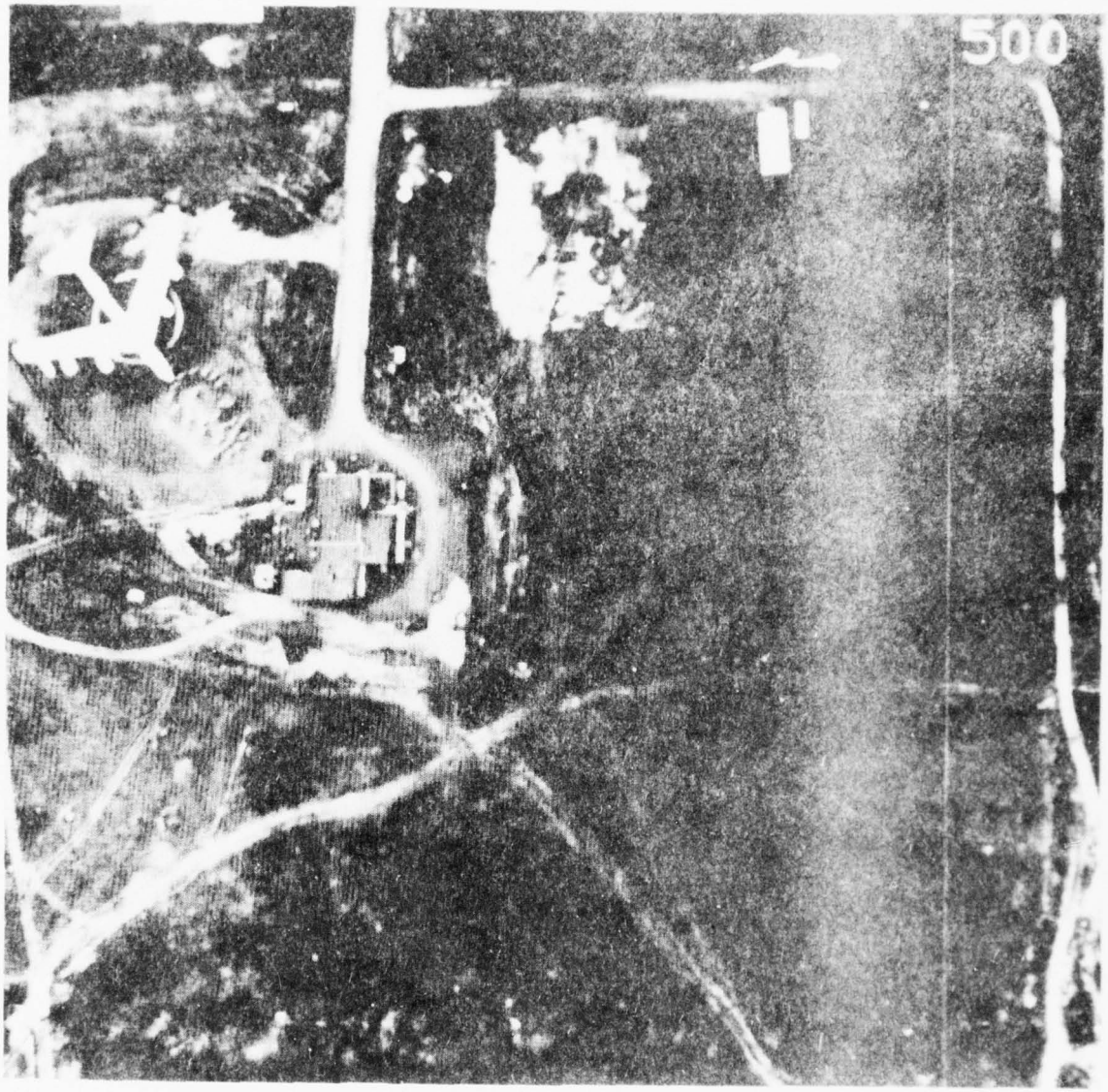


Fig. V-1. The original image "SAM3" used in the photo analyst rankings. Size 512 x 512 at 8 bits/pixel. Photo number 500.



Fig. V-2 The original image "AP2" used in the photo analyst rankings. Size 512 x 512 at 8 bits/pixel. Photo number 600.

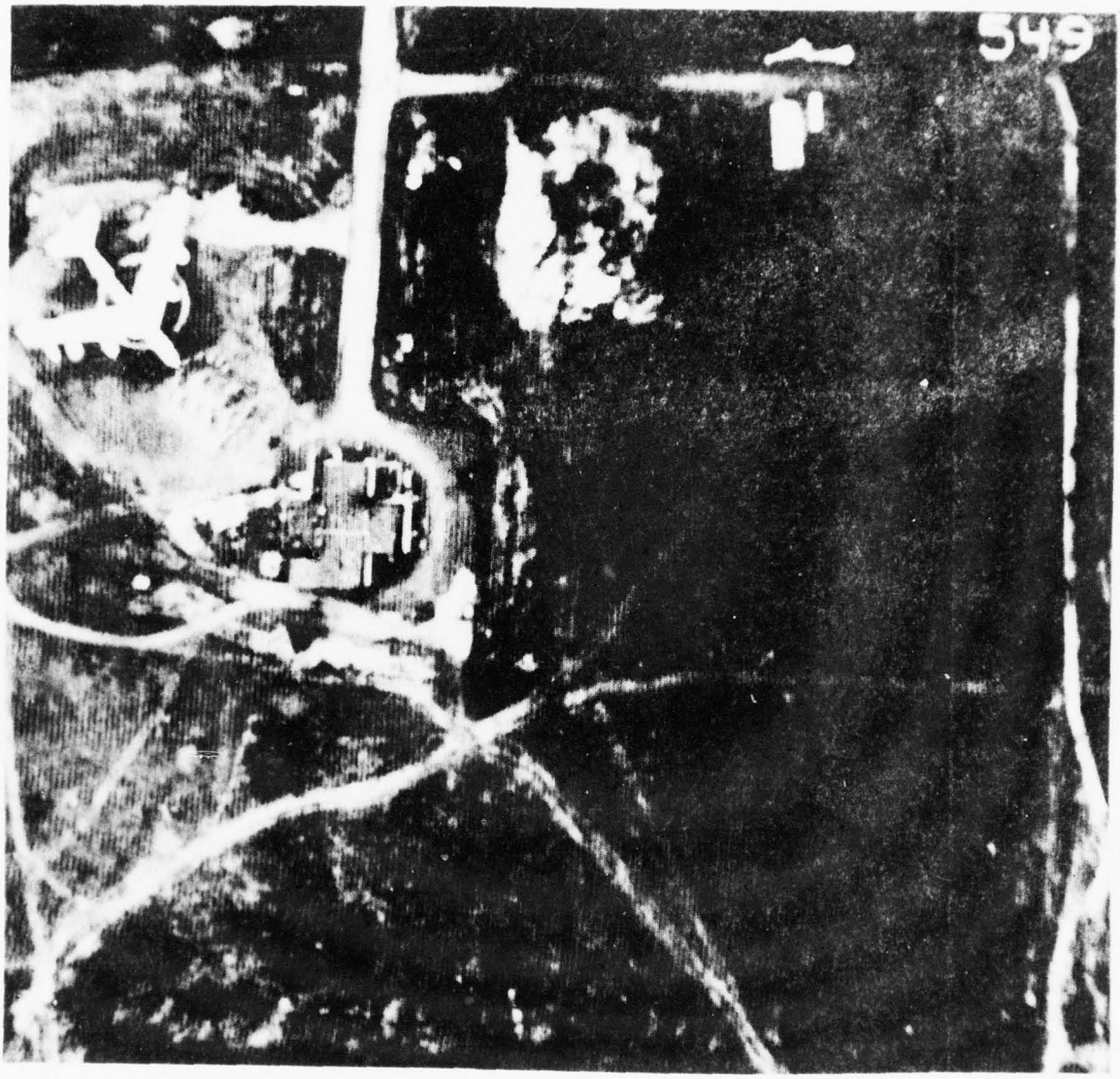


Fig. V-3 The frequency-domain four class zonal techqiue as applied to SAM3: 1.6 bits/pixel, Photo number 549.

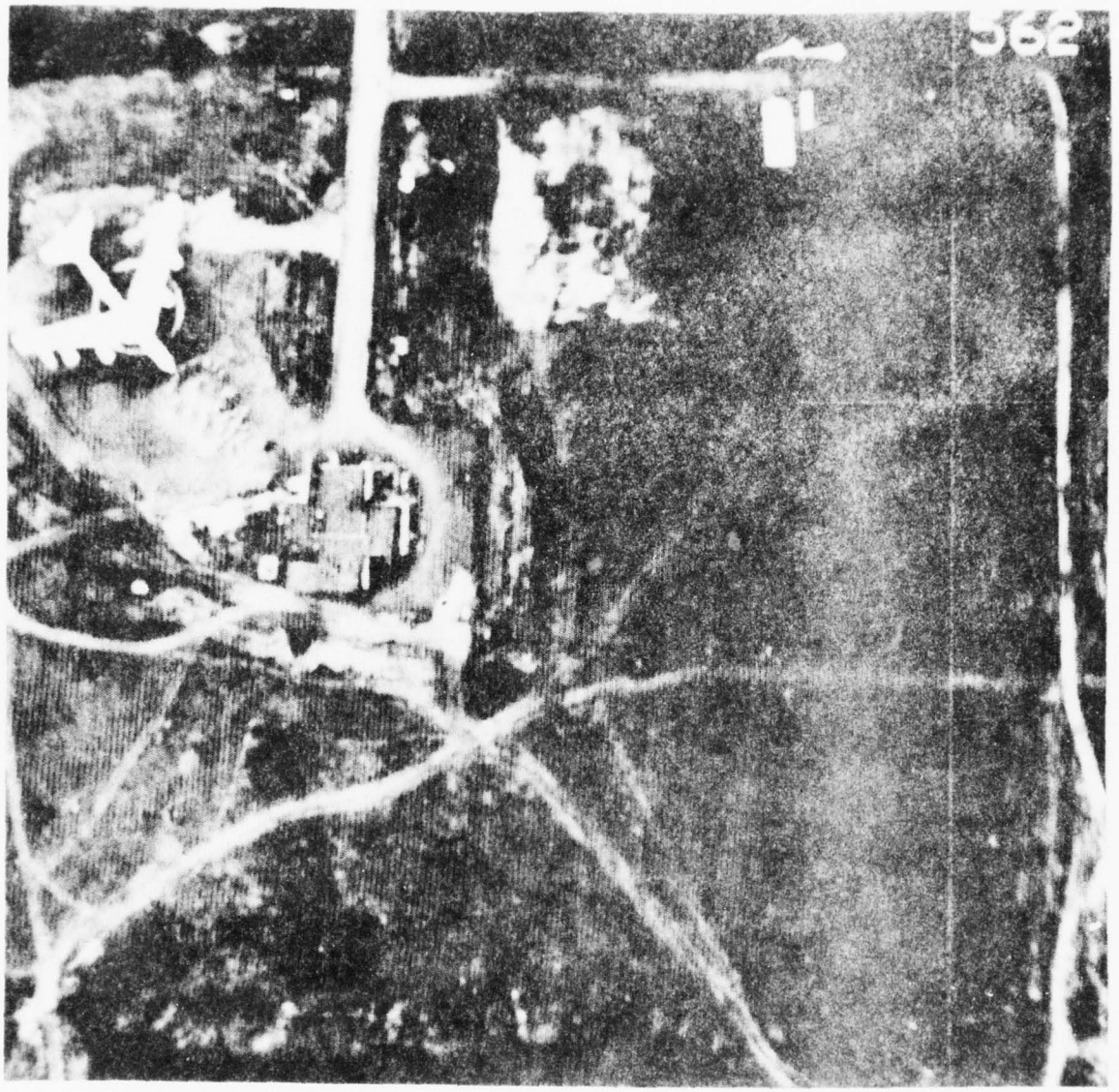


Fig. V-4 The frequency-domain four class zonal technique (with spatial domain preprocessing) as applied to SAM3 - 1.4 bits/pixel. Photo number 562



Fig. V-5 The hybrid method applied to SAM3: 1.6 bits/pixel. Photo number 514.

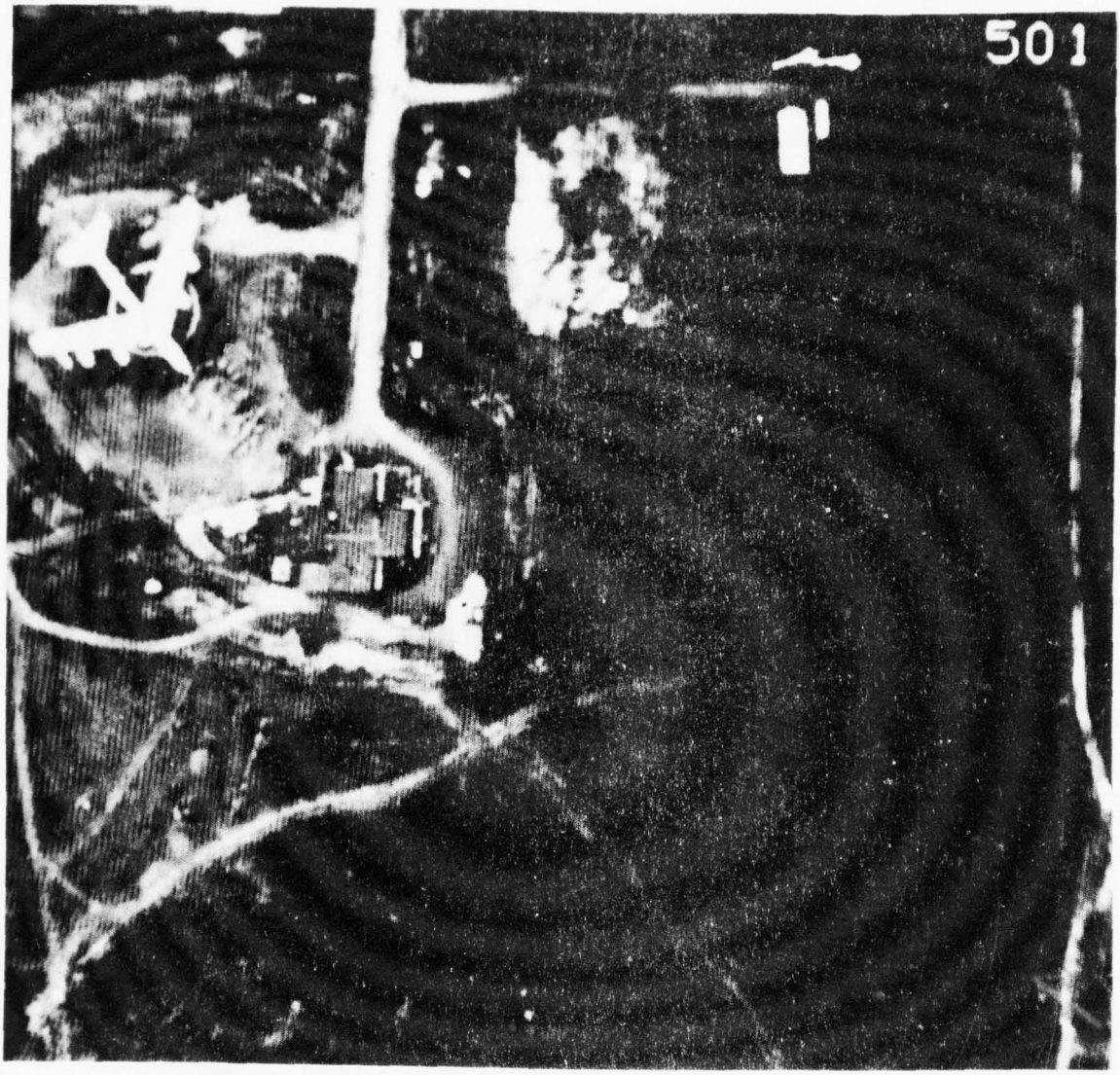


Fig. V-6 The moment-preserving block truncation method applied to SAM3: 1.6 bits/pixel. Photo number 501.

V. Test Results

V.1. Original Images and Reconstructed Results

The two 512 x 512 images subjected to the various coding algorithms described in this report are shown in Figures V-1 and V-2. Both of these images were originally quantized at 8 bits per pixel. The first image, "AP2", was chosen due to its wide variability of image characteristics. Of interest are the city areas, the airport runways, and the boats and other small objects in the water. The second original, "SAM3" is part of the Northeast Test Site Area.

Shown in Figures V-3 through V-6 are several decoded versions of the "SAM3" original when processed by four of the techniques described in Sections II, III, and IV. As discussed in Sub-section II.3, Figures V-3 and V-4 are results obtained from the same basic coding algorithm. However, Figure V-4 includes spatial domain pre-processing. Figures V-5 and V-6 reflect results of applying the "hybrid" and the "moment preserving block truncation" technique to the same SAM3 image. In subjective rankings, the "hybrid" method was usually rated better than the "block truncation" technique. However, both were typically rated lower than the four class methods. On the other hand, it should be born in mind that both of these algorithms represent much less of a computational burden than the zonal methods.

In Figures V-7 through V-10 we make a comparison between a spatial domain method, namely the basic block truncation algorithm and the four class zonal method based on spatial domain criteria. See sub-section II.2.3. The block truncation results of Figure V-7 were ranked higher by the photoanalysts than the four class results of Figure V-8. The original image in both of these cases was "AP2" sampled at a size of 512 x 512. When

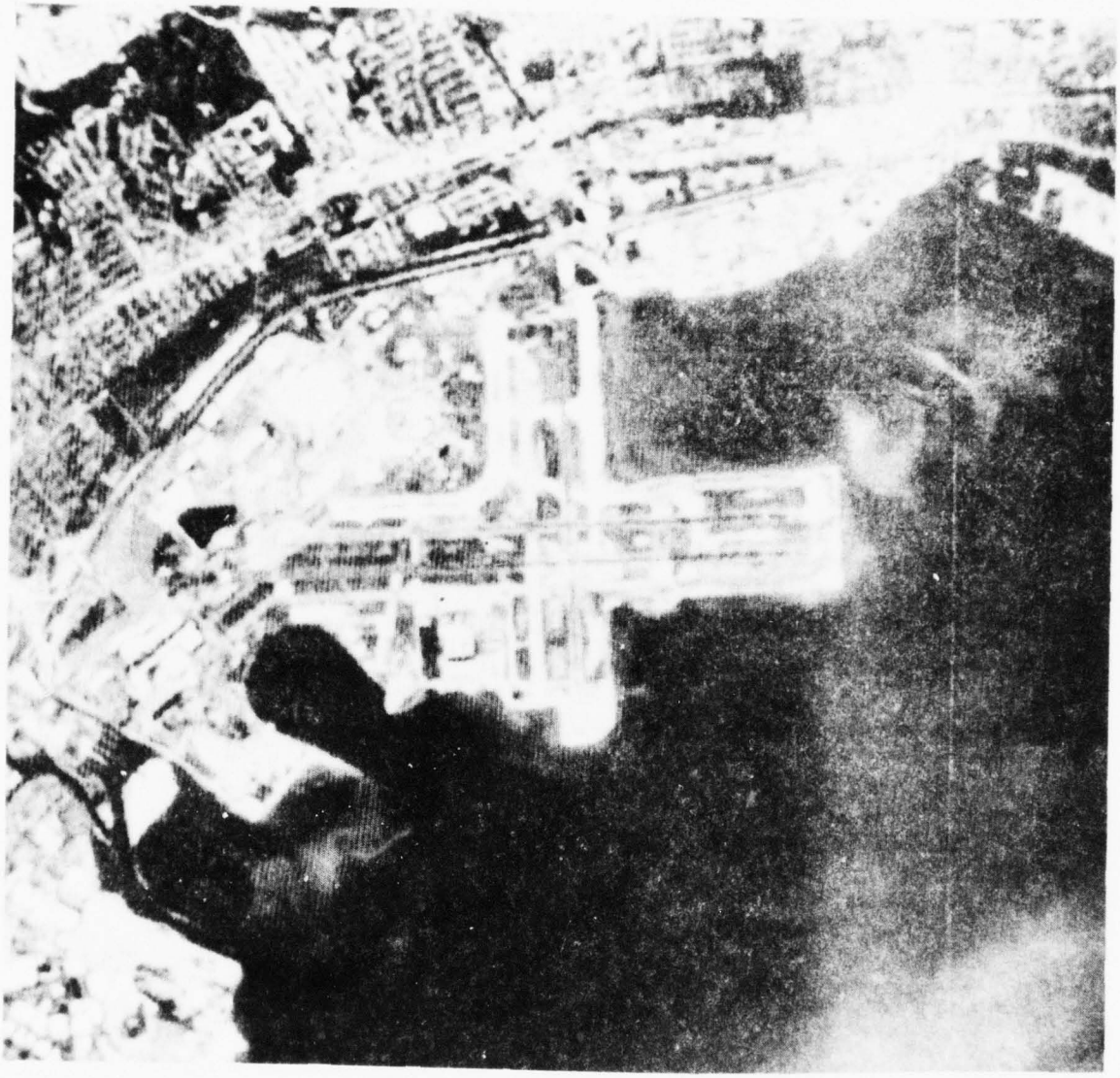


Fig. V-7 The basic block truncation algorithm applied to AF2: 1.6 bits/pixel. Photo number 631.



Fig. V-8 The spatial-domain four class zonal technique (using unequal classes) as applied to AP2: 1.6 bits/pixel. Photo number 643.

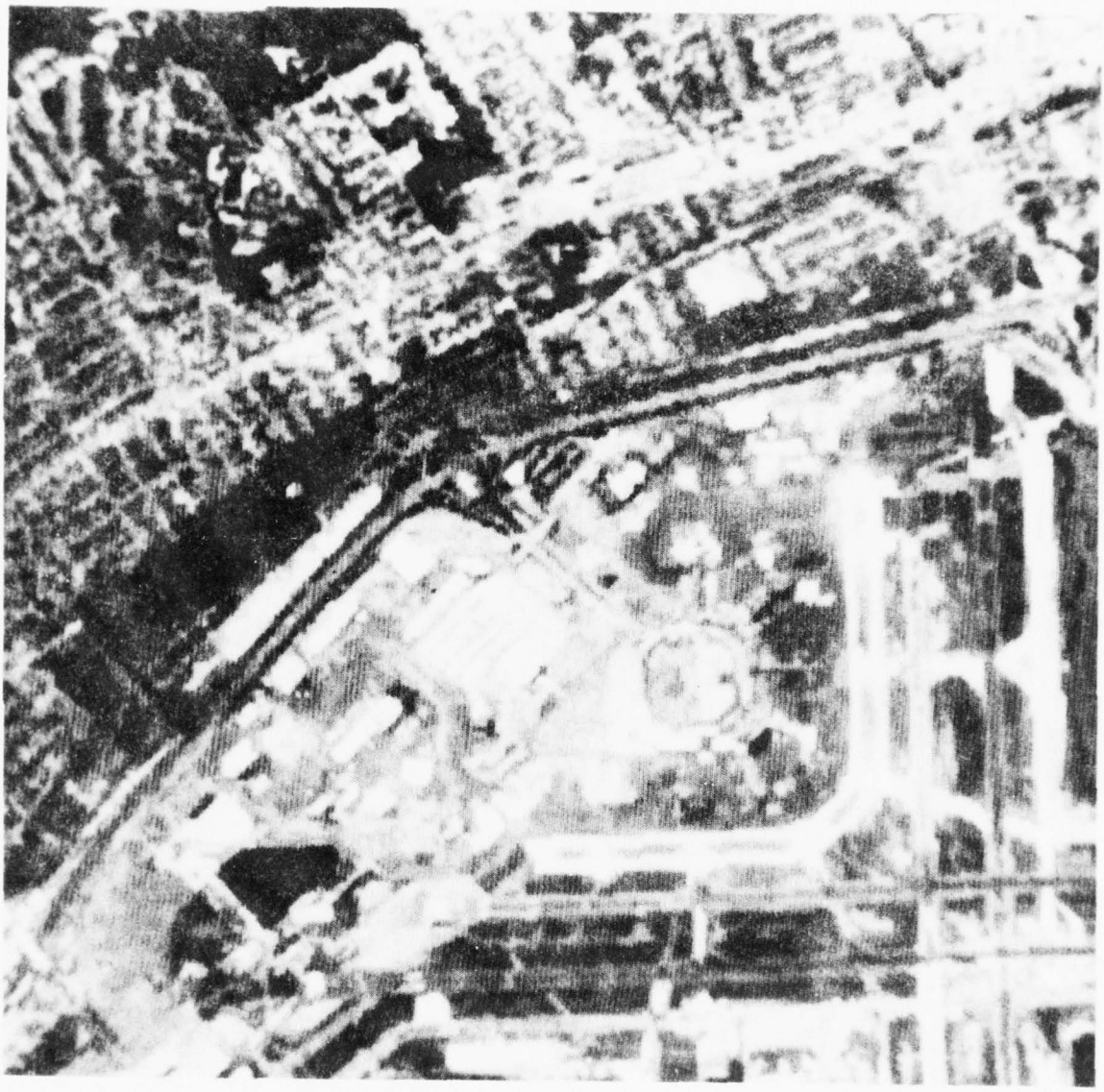


Fig. V-9 The basic block truncation method applied to a 256 x 256 section of AP2: 1.6 bits/pixel. Photo number 407.

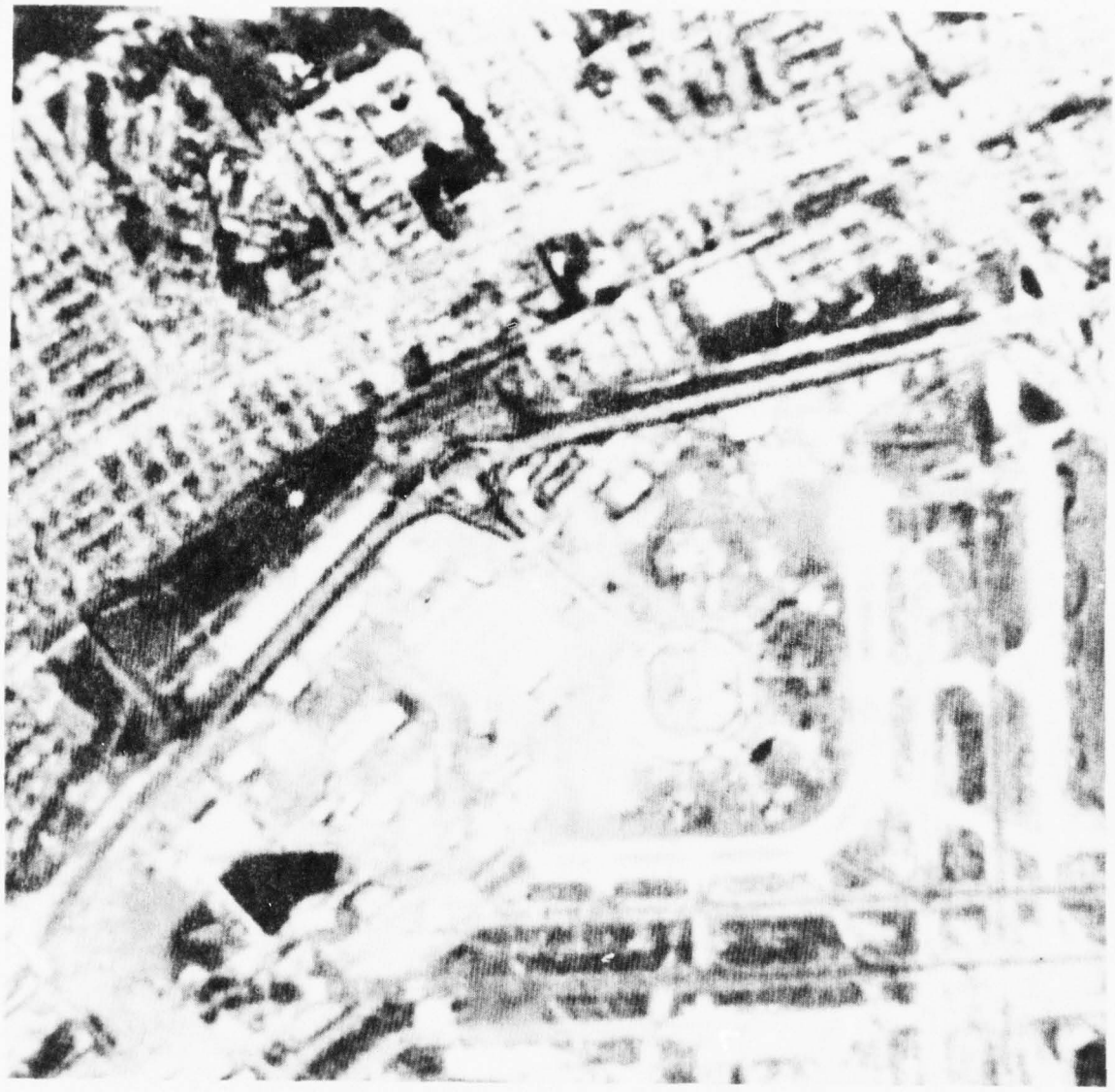


Fig. V-10 The spatial-domain four class zonal technique (using equal classes) as applied to a 256 x 256 section of AP2: 1.6 bits/pixel. Photo number 405.

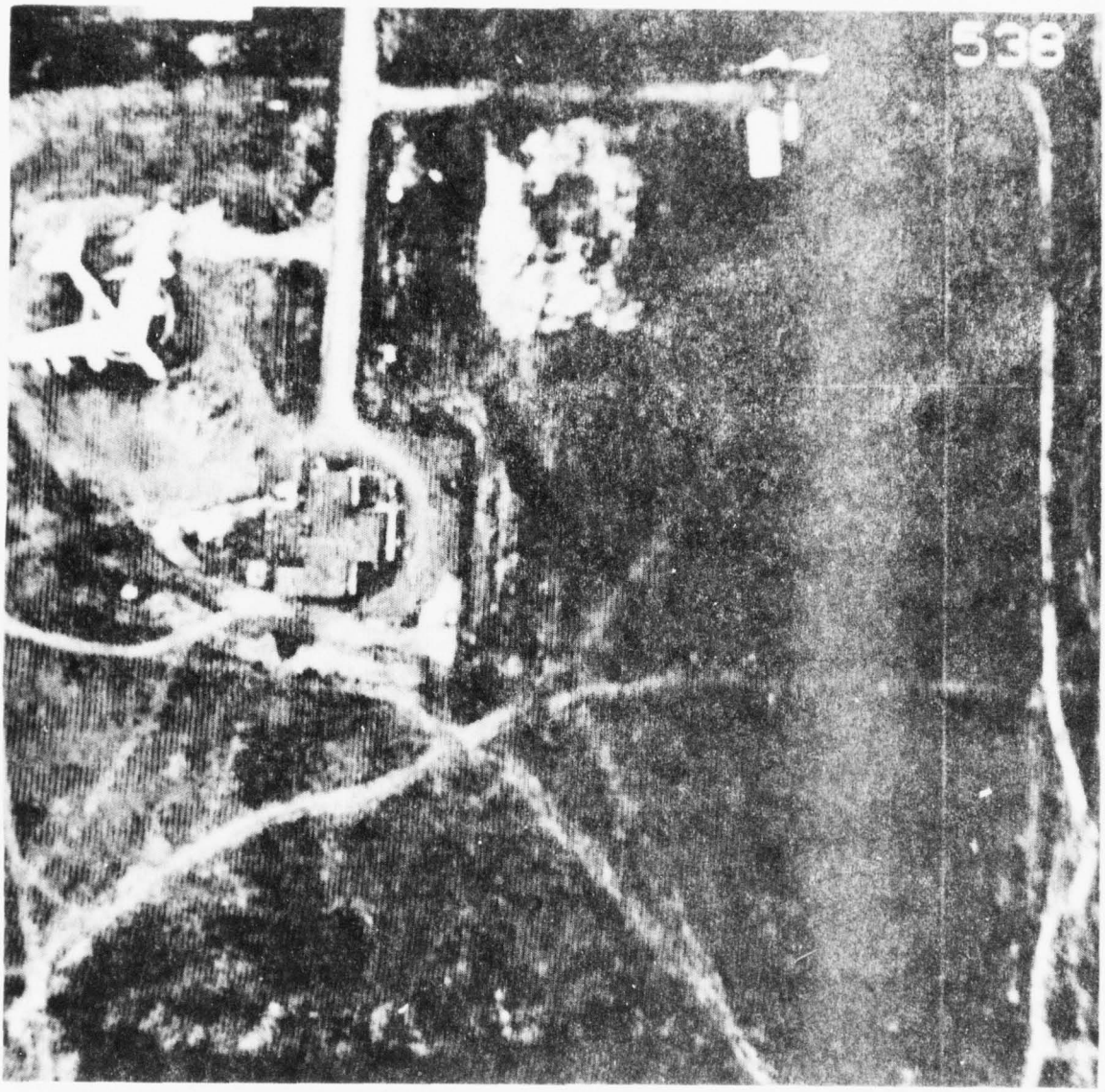


Fig. V-11 The frequency-domain four class zonal method as applied to SAM3: 0.5 bits/pixel. Photo number 538.

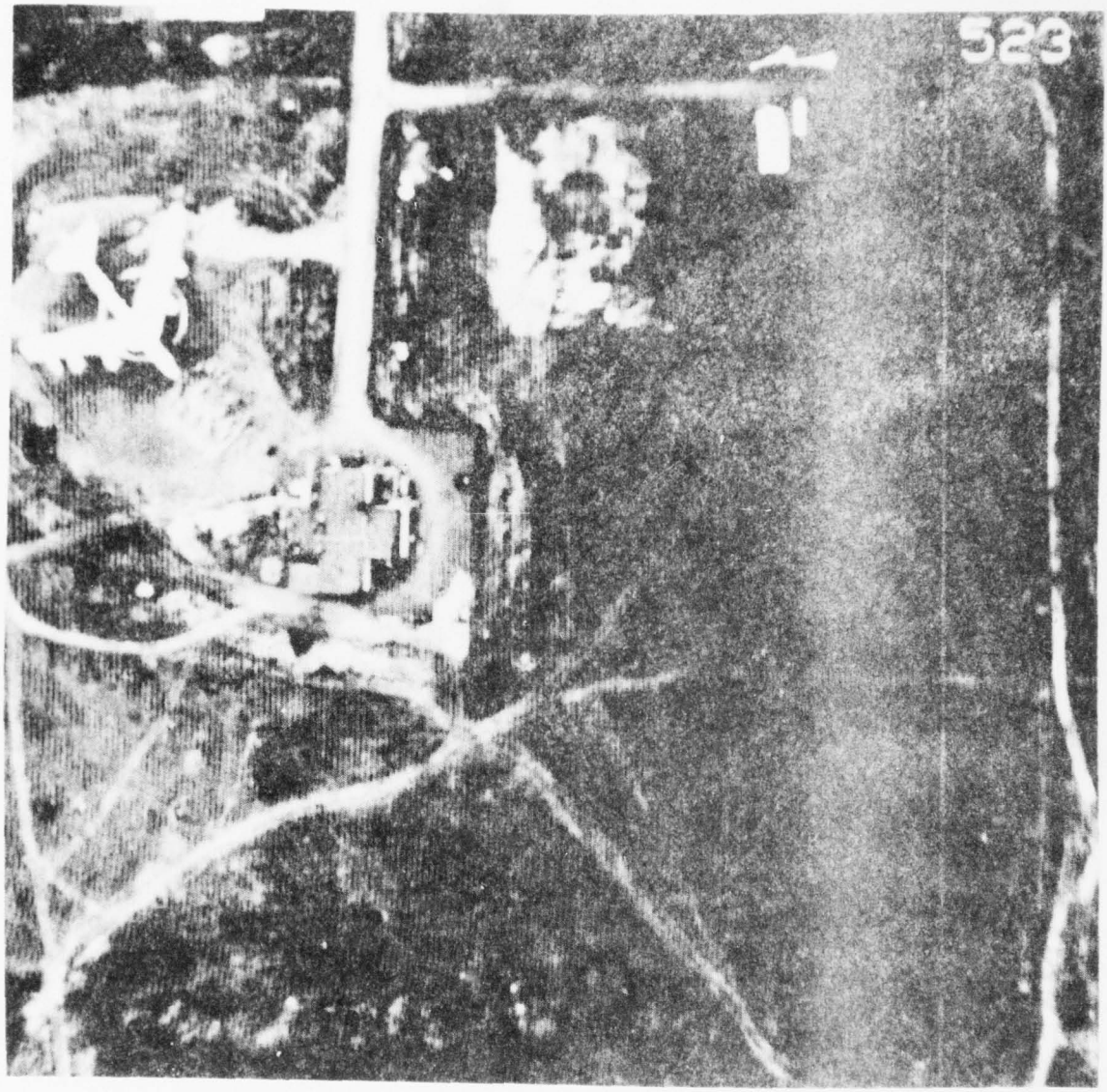


Fig. V-12 The spatial-domain four class zonal method as applied to SAM3 with a channel bit error rate of 10^{-3} : 1.6 bits/pixel. Photo number 523.



Fig. V-13 The hybrid coder in application to AP2 with a channel bit error rate of 10^{-3}: 1.6 bits/pixel.

a small section of this image, of size 256 x 256, is blown up for evaluation, the results of Figures V-9 and V-10 are obtained. It is interesting to note that the expansion of these smaller images resulted in a relative reversal of the photo-interpreters evaluation of these two techniques. The basic block truncation method is now ranked significantly lower than the spatial domain four class methods.

In order to compare the quality degradation experienced when passing from 1.6 bits per pixel to a compression rate of 0.5 bits per pixel, we include Figure V-11. This is the four class frequency domain technique discussed in Sub-section II.2.2. The image is once again SAM3. The effects of channel errors on two of the techniques studied in our work are depicted in Figures V-12 and V-13. The coding method used in Figure V-12 is the spatial domain four class method. The hybrid system generated the noisy channel performance shown in Figure V-13. The error rate used in both of these tests was 10^{-3}.

V.2. Ranking Results

The photo analysts were supplied a ranking form for each of the 7 photo sets. One sample form appears in [1, p.98]. In Table V-2 appear the ranking results. Abbreviations used in this table are defined in Table V-1. Notice that each of the seven sets was ranked by five analysts. In each case a ranking of 1 denoted the best (in that analyst's opinion) reproduction of the original. Space was provided on the form for additional comments such as excellent (EX) and unacceptable (X). The presence of either of these two comments appears in the raw data of Table V-2 as a superscript to the corresponding ranking number. The right three columns of each set list the "averaging ranking", mean square, and mean absolute reconstruction error for each method. The average ranking was calculated by dropping the

4ZCS	-	four class zone method first presented by Chen and Smith [2].
4ZFC	-	four class zone method using feature clustering described in Sec. II.2.2.
4ZFCR	-	same as 4ZFC but with pre and post processing as described in Sec. II.3.
4ZSF	-	four class zone using spatial criteria and equal size classes described in Sec. II.2.3.
4Z5V	-	four class zone using spatial criteria and variable size classes described in Sec. II.2.3.
HYB	-	hybrid as described in Sec. III.
BTC	-	basic block truncation algorithm as described in Sec. IV.
BTCMP	-	BTC with moment preserving threshold as discussed in Sec. IV.
BTCH	-	BTC with hybrid addition (1.8 bits/pixel total) as discussed in Sec. IV.
BTCMS	-	BTC with minimum mse (2.0 bits/pixel) as discussed in Sec. IV.
BTCMA	-	BTC with minimum mae (2.0 bits/pixel) as discussed in Sec. IV.

Abbreviations used in ranking results presented in Table V-2.

TABLE V-1

TABLE V-2

"Raw data" as taken from analysts forms. Also shown for each method are an average ranking and the computed mean-square error and mean absolute error for that method. The average rank was calculated by dropping the highest and lowest rankings and averaging the remaining three. The method name abbreviations are described in Table V-1.

(cont. on next page)

SAM3, 1.6 bits/pixel, 512x512, no errors

Method	Photo No.	Analyst #1	#2	#3	#4	#5	AVG RANK	MSE	MAE
4ZFCP	562	2	1	1^{EX}	2^{EX}	7	1.7	31.60	3.80
4ZSV	512	5	2	3	1^{EX}	2^{EX}	2.3	31.55	3.89
4ZFC	549	1^{EX}	5	2	3^{EX}	6	3.3	28.71	3.74
4ZSF	587	3	6	4	8	5	5.0	34.26	3.91
BTCH	552	6	7	5	6	3^{EX}	5.7	48.43	4.53
BTCMP	501	7	4	8	7	1^{EX}	6.0	50.13	4.69
4ZCS	517	4	8	6	5^{EX}	8	6.3	42.79	4.43
HYB	514	9	9^{X}	9^{X}	4^{EX}	4	7.3	94.37	5.10
BTC	576	8	3	7	9	9	8.0	53.56	4.70

Correlation Coefficients: Avg. rank - mse $\rho = 0.71$

Avg. rank - mae $\rho = 0.87$

$SAM3_{ex}$, 1.6 bits/pixel, 256x256, no errors

4ZFC	453	2	1	1^{EX}	2	3	1.7
4ZFCP	451	1	2	3	2^{EX}	6	2.0
4Z5V	457	4	5	2	3	1^{EX}	3.0
4ZCS	454	3	3	4	5	4	3.7
4ZSF	456	5	4	5	4	2	4.3
HYB	452	6	6^{X}	6^{X}	6^{X}	6^{X}	6.0

$SAM3_{ex}$, 1.6 bits/pixel, 256x256, no errors

BTCH	459	1^{X}	2	1	1	1	1.0
BTCMA	462	2^{X}	1	4	3	2	2.3
BTCMS	461	3^{X}	5	2	2	3	2.7
BTCMP	458	4^{X}	4	3	5	4	4.0
BTC	460	5^{X}	3	5^{X}	4	5	4.7

Table V-2 (cont.)

SAM3, 1.6 bits/pixel, 512x512, 10^{-3} error prob.

		Analyst					AVG		
Method	Photo No.	#1	#2	#3	#4	#5	RANK	MSE	MAE
4ZSF	523	1	2	1	1	1^{EX}	1.0	45.67	4.27
4ZCS	568	2	1	4	3	2^{EX}	2.3	75.40	4.99
4ZFC	503	3	4	2	2	6	3.0	44.17	4.20
HYB	590	6	3	3	6	3	4.0	103.06	5.47
4ZSV	539	4	5^{X}	5^{X}	4	5	4.7	65.54	4.47
BTCP	573	5	6^{X}	6^{X}	5	4	5.3	93.11	5.14

Correlation Coefficients: Avg. rank - mse $\rho = 0.62$

Avg. rank - mae $\rho = 0.46$

SAM3, 0.5 bits/pixel, 512x512, no errors

4ZFC	538	2	1	5	1	1^{EX}	1.3	81.61	6.48
4ZFCP	557	1	5	1^{EX}	2	2	1.7	86.21	6.35
4ZSF	592	3	3	3	4	4	3.3	94.60	6.89
4ZCS	542	4	4	2	5	3	3.7	107.70	7.37
4ZSV	554	5	2	4	3	5	4.0	96.24	6.96
HYB	577	6^{X}	6^{X}	6^{X}	6^{X}	6^{X}	6.0	228.18	8.74

Correlation Coefficients: Avg. rank - mse $\rho = 0.85$

Avg. rank - mae $\rho = 0.93$

Table V-2 (cont.)

AP2, 1.6 bits/pixel, 512x512, no errors

		Analyst					AVG		
Method	Photo No.	#1	#2	#3	#4	#5	RANK	MSE	MAE
BTC	631	1^{EX}	5	6	3^{EX}	1^{EX}	3.0	76.66	5.70
4ZFCP	672	2^{EX}	3	4	2^{EX}	6	3.0	30.81	3.89
4ZSF	643	6^{EX}	4	1	1^{EX}	5	3.3	24.74	3.58
HYB	636	3^{EX}	6	3	4^{EX}	3^{EX}	3.3	47.76	4.82
BTCMP	627	5^{EX}	1	7	6	2^{EX}	4.3	78.14	5.74
4ZFC	615	4^{EX}	7	2	5	4	4.3	25.20	3.58
4ZSV	606	7^{EX}	2	5	7	7	6.3	36.75	4.44

Correlation Coefficients: Avg. rank - mse $\rho = -0.12$

Avg. rank - mae $\rho = -0.03$

$AP2_{ex}$, 1.6 bits/pixel, 256x256, no errors

4ZFCP	402	1^{EX}	2	1^{EX}	4	2^{EX}	1.7
4ZSF	404	3	1	2	3^{EX}	4	2.7
4ZSV	405	2	4	3	5	1^{EX}	3.0
4ZFC	403	4	3	4	2^{EX}	3^{EX}	3.3
HYB	401	5	7^{X}	5	1^{EX}	5	5.0
BTCMP	406	6^{X}	5	6^{X}	6	6	6.0
BTC	407	7^{X}	6	7^{X}	7	7	7.0

highest and lowest ranking given each method and averaging the remaining three.

Table V-2 also lists correlation coefficients relating the interpreter's average ranking of a reconstructed image to its calculated m.s.e. and m.a.e. (mean absolute error). These were calculated as

$$\hat{\rho}_{xy} = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y}$$

where $\hat{\sigma}_x$ is the sample standard deviation of the average interpreter rankings (across all reconstructed images within a set), while $\hat{\sigma}_y$ is the sample standard deviation of either the m.s.e. or m.a.e. across that same set. And $\hat{\sigma}_{xy}$ is the sample cross correlation. The sampling bias effects were not taken into consideration.

V.3. Discussion

The average rakings in general are fairly well correlated with mean square error and mean absolute error. However, the variation among individual analysts' rankings show the subjectivity of such a test procedure and/or similarities of many of the reconstructed results. In a few cases, however, the subjective rankings differed from the m.s.e. rankings. This is most noticeable in the case of AP2 where the correlation coefficient is actually negative. In this particular set, the reconstructed picture quality was high enough that the analysts could not agree on similar rankings.

The four class zone methods performed the best in terms of subjective rankings and mean square error. At 0.5 bits per pixel, they were the only acceptable methods and the feature clustering method of bit assignment [Fig. V-11] was ranked the best by the photoanalysts. Although some degradation is visible in this image, it received an "excellent" by one of the analysts.

At 1.6 bits per pixel, all four class zone methods performed well. The preprocessing to improve performance in the dark regions (Sec.II.3) resulted in the best overall rankings, thus showing the viability of such a method as applied to aerial reconnaissance imagery. In terms of computational load and performance the preprocessed feature clustering (4ZFCP) method appears presently to produce the best overall results.

The hybrid method is now a viable moderate computational alternative to the two-dimensional transform methods even at error rates up to 10^{-3} when coded at 1.6 bits per pixel. Although it was ranked below the four-class zone methods, it generally provided acceptable results. The computations involved in the hybrid approach are significantly less than those required by the two-dimensional transform methods. More importantly, the number of mass storage picture accesses required by hybrid coding is two, instead of the three required by the 2-D methods.

The best performing spatial technique that we have tested is block truncation coding. The overall rankings of BTC are generally below those of the transform techniques. However, under normal viewing conditions (individual pixels not visible as separate entities), BTC provided acceptable and at times, excellent, results. This implies that post processing at the receiver would improve subjective performance when individual pixels are visible. The computational savings in using this technique are enormous. This is the only real time, single image access technique rated acceptable here and in [1,16].

VI. Conclusions and Future Research Directions

Over the past 1-1/2 years, our research has indicated that for good subjective reproduction of high resolution aerial imagery, compression ratios in the range of 0.5 to 1.5 bits/pixel are achievable using adaptive

two-dimensional block transform techniques. Random channel errors of 10^{-4} are insignificant and of 10^{-3} are tolerable.

Between 1.0 and 2.0 bits/pixel hybrid coding becomes almost competitive in quality and offers computational advantages.

Above 1.5 bits/pixel, several spatial coding possibilities exist which offer tremendous computational and storage advantages as well as respectable, if not excellent, performance.

We feel that three major areas of study should be investigated to make the application of these coding methods more desirable and easily implemented. These are:

(1) Efficiency Improvements - The best performing coding methods (2-D transforms) are the most computationally intensive. The best method requires 3 passes through the data to categorize each 16 x 16 block, collect block statistics and to do bit assignments and coding. Upon observing the bit assignments and categorizations, we feel that some standard bit assignments and categories could be derived and the 2-D transform methods be made much more efficient by assigning each block to one of these predetermined categories and sending a short code to the receiver for each indicating the assigned category. This would eliminate all picture storage or rescanning requirements in that each block could be processed and transmitted independent of the others. This also reduces the transmitted overhead because the bit maps can be prestored at the receiver.

(2) Technique Combinations - Coding methods are usually quite data dependent in regard to subjective performance. For example, pic-

ture blocks consisting of one single bright pixel and many dark ones are most efficiently coded using a spatial technique, while more gradual variations are more suited for transform coding. We have adapted all of our techniques to operate on 16 x 16 blocks so that the methods can be mixed to provide an optimum strategy. This method is actually an extension of the efficiency method suggested in part (1) where the various categories not only include various bit assignments but also can indicate various coding methods. The category selection would involve frequency domain (how much energy in each region) and spatial domain (presence of spots or edges) measurements.

(3) Source Error Correction - It has been mentioned that the presence of uncorrected channel errors can cause significant distortion in the received picture. The need for bandwidth compression eliminates the possibility of using additional bits for channel error detection and correction. However, it is possible in many cases for the receiver to find source errors and correct them. A simple method would be to check the boundary between a block to be tested and its neighbors. A discontinuity in all (or most) boundary points indicates a bad block and the receiver could then test the dc and other low frequency coefficients to try to locate the error. We feel that for several of our coding methods implemented, such a system would give an order of magnitude receiver performance improvement when operating in the presence of channel errors.

In addition to the general areas outlined above, three specific techniques that have the potential for development are now described:

Mixed Basis Transforms

In matrix form, an orthogonal transform of an NxN image block $\underset{\sim}{G}$ can be expressed as

$$\mathcal{T}\{\underset{\sim}{G}\} = \underset{\sim}{U}^T \underset{\sim}{G} \underset{\sim}{U}$$

where a superscript T denotes the matrix transpose operation while each column of $\underset{\sim}{U}$ holds samples of a basis function of the transform type being used (Cosine, Fourier, Slant, etc.). It is known that since the above equation can be expressed as

$$\mathcal{T}\{\underset{\sim}{G}\} = \underset{\sim}{U}^T \underset{\sim}{G} \underset{\sim}{U} = [(\underset{\sim}{G}\underset{\sim}{U})^T \cdot \underset{\sim}{U}]^T$$

the two-dimensional transform operation may be viewed as a three-step process: 1.) "U-transform" each row of $\underset{\sim}{G}$ separately, replacing that row with its one-dimensional transform; 2.) transpose this new matrix of one-dimensional transforms. The i^{th} row of the resulting matrix now contains all the i^{th} basis function weights from row zero of the original $\underset{\sim}{G}$ matrix down through row N-1. (If the rows of $\underset{\sim}{G}$ are highly correlated, the rows of $(\underset{\sim}{G}\underset{\sim}{U})^T$ will be slowly-varying from column zero through column N-1.); 3.) "U-transform" the rows of $(\underset{\sim}{G}\underset{\sim}{U})^T$, again treating these as separate one-dimensional objects.

There is no theoretical or computational reason why the transforms used in steps 1.) and 3.) need be of the same type [13]. Though it is rarely done, the "row" transform of step 1.) can be of type $\underset{\sim}{U}$ while the column transform of step 3.) can be of type, say, $\underset{\sim}{V}$. Thus we have a mixed (or hybrid) transform method expressible as

$$\{\underset{\sim}{G}\} \triangleq \underset{\sim}{V}^T \underset{\sim}{G} \underset{\sim}{U}$$

Indeed, the hybrid coding method reported on in Sec. III (and in [1]) does something like this by replacing the V-transform operation with a DPCM coder.

Since many image blocks will exhibit the high correlation (or slowly - varying nature), mentioned in connection with step 2.). Above, it should be worthwhile to investigate alternate "V-transforms" that might compact such slowly-varying "row signals" more effectively than a repeated application of $\underset{\sim}{U}$. (In all our present cases, $\underset{\sim}{U}$ is the cosine transform.) Of most interest at this early stage would be the slant [14] transform since it affords a linearly graded basis function; but others should also be tried, as should other spatial domain interpolator besides DPCM.

MAPS Improvements

Of those coding algorithms we have studied, the MAPS [15] approach is the most computationally efficient. Hence, we propose to examine this method more closely, hopefully to improve its rated performance in photo reconnaissance work.

The two fundamental aspects of MAPS coding are 1.) its pixel ordering sequence which, as we have shown in [1, p.45], allows the coarsely-variable length MAPS records to be transmitted at error rates as high as 10^{-3} without (usually) the loss of transmitter-receiver synchronization; and 2.) the representation of a spatial group of pixels by a single number: the group intensity mean. It is this latter aspect we hope to improve.

Within any local area of an image, the compression rates at which a MAPS coder may operate are few in number and of progressively wider spacing. See Table VI-1. The result of such a coarse partition of the overall compression range must frequently give rise to a "feast or famine" situation

in which one compression rate affords far more resolution than desired, while the next higher rate yields unacceptably poor performance. However, if, after each line having an asterisk in Table VI-1, we insert an additional mode of compression operation, we obtain the somewhat finer partition of Table VI-2. In each of these new cases we have provided for <u>two</u> (rather than one or four) words of intensity information: One 6-bit word (as before) and one (new) 8-bit word. This latter quantity could be split into subfields as suggested below.

Some quantities for which this new 8-bit field could be reserved are:

1.) A variance, which, together with the mean, could establish a random number generator which would define pixel intensities within a single block at the receiver. This should be effective in removing the block edge artifacts currently apparent in MAPS-coded images. (Something like this was suggested in [15], though no separate code word was to be reserved for it.) This variance quantity would be in some way dependent on 4.) the block size, b.) the pixel intensity variance within the block, or c.) the intensity variances at the block edges. If no multiplications are desired, the square root of the variance (s.d.) could be replaced by a mean absolute value.

2.) The 6 and 8-bit work could be jointly subdivided in such a way as to allow the transmission of the a.) center value, b.) slope, and c.) coarse slope direction of an interpolating intensity plane. Such a plane should be a better pixel interpolator than a single d.c. level.

AD-A071 088 PURDUE UNIV LAFAYETTE IND SCHOOL OF ELECTRICAL ENGI--ETC F/G 9/4
IMPROVEMENTS IN SOME IMAGE COMPRESSION TECHNIQUES FOR AERIAL RE--ETC(U)
MAY 79 O R MITCHELL, S C BASS, E J DELP F30602-78-C-0102
UNCLASSIFIED RADC-TR-79-99 NL

2 OF 2
AD
A071 088



END
DATE
FILMED
8 -79
DDC

	Merged Group Size	Intensities in Group	Bits Transmitted (Including Overhead)	Effective Compression Rate
*	2 x 2	4	30.0937	7.5 bits/pixel
	2 x 2	1	12.0937	3.02
*	4 x 4	4	30.375	1.897
	4 x 4	1	12.375	.773
*	8 x 8	4	31.5	.4923
	8 x 8	1	13.5	.2109
*	16 x 16	4	36.0	.1405
	16 x 16	1	18.0	.0703

Table VI-1: The instantaneous coding rates available to the MAPS algorithm as implemented at Purdue.

	Merged Group Size	Intensities in Group	Bits Transmitted (Including Overhead)	Effective Compression Rate
	2 x 2	4	30.0937	7.5 bits/pixel
*	2 x 2	2	20.0937	5.02
	2 x 2	1	12.0937	3.02
	4 x 4	4	30.375	1.897
*	4 x 4	2	20.375	1.275
	4 x 4	1	12.375	.773
	8 x 8	4	31.5	.4923
*	8 x 8	2	21.5	.3360
	8 x 8	1	13.5	.2109
	16 x 16	4	36.0	.1405
*	16 x 16	2	26.0	.1016
	16 x 16	1	18.0	.0703

Table VI-2: Inserting new MAPS coding rate levels (marked with an asterisk) results in a somewhat finer partition for better adaptive coding.

3.) The 8-bit word could be split into two subfields to allow the specification of a.) an interpolating basis function aplitude, and b.) which basis function. For the latter, three bits could select one of the eight low frequency 2-D cosine transform basis functions nearest d.c.

BTC Improvements

We have recently been experimenting with both pre- and post-processing of the image relative to BTC. The post-processing has significantly reduced the m.s.e. of the reconstructed image and we feel the subjective performance will improve. Much of the visible error in BTC is sharp edge introduction due to the one bit quantization. The post-processing takes the form of an optimum estimation procedure at the receiver based on picture and artifact statistics. The pre-processing we have done is that of selective blurring to enhance the threshold selection. We feel both of these methods should be explored to improve the performance of BTC.

References

1. O.R. Mitchell, S.C. Bass, E.J. Delp, and T.W. Goeddel, "Coding of Aerial Reconnaissance Images for Transmission Over Noisy Channels," RADC-TR-78-210, Rome Air Development Center, Griffiss Air Force Base, N.Y., September 1978, A061539.

2. W.H. Chen and C.H. Smith, "Adaptive Coding of Monochrome and Color Images," IEEE Transactions on Communications, Vol. COM-25, November 1977, pp. 1285-1292.

3. N. Ahmed, T. Natarjan, and K.R. Rao, "Discrete Cosine Transform," IEEE Transactions on Computers, Vol. C-23, January 1974, pp. 90-93.

4. W.H. Chen, C.H. Smith, and S. Fralick, "A Fast Computational Algorithm for the Discrete Cosine Transform," IEEE Transactions on Communications, Vol. COM-25, September 1977, pp. 1004-1009.

5. L.D. Davisson, "Rate-Distortion Theory and Application," Proceedings of IEEE, Vol. 60, July 1972, pp. 800-828.

6. J.L. Mannos and D.J. Sakrison, "The Effects of a Visual Fidelity Criterion on the Encoding of Images," IEEE Trans. on Info. Theory, Vol. IT-20, pp. 525-536, July 1974.

7. A. Habibi, "Hybrid Coding of Picture Data," IEEE Transactions On Communications, Volume COM-22, pp. 614-624, May 1974.

8. N. Ahmed and K. Rao, Orthogonal Transforms For Digital Signal Processing, Springer-Verlag: Berlin, 1975.

9. A. Papoulis, Probability, Random Variables, and Stochastic Processes, McGraw-Hill: New York, 1965.

10. J. Max, "Quantizing for Minimum Distortion", IRE Transactions On Information Theory, Volume IT-6, pp. 7-12, March 1960.

11. E.J. Delp and O.R. Mitchell, "Spatial Image Compression Using Block Truncation Coding," submitted to IEEE Trans. on Communications.

12. E.J. Delp and O.R. Mitchell, "Some Aspects of Moment Perserving Quantizers," submitted for presentation at 1979 IEEE International Conference on Communications, Boston, Mass.

13. T.S. Huang, ed., Picture Processing and Digital Filtering, Springer-Verlag, New York, 1975.

14. W.K. Pratt, Digital Image Processing, John Wiley and Sons, New York, 1978.

15. A.E. Labonte and C.J. McCallum, "Image Compression Techniques," Rome Air Development Center, Tech. Rept. No. RADC-TR-77-405, December 1977. A050679.

16. W.L. Eversole, D.J. Mayer, F.B. Fruzee, and T.F. Cheek, "Investigation of VLSI Technologies for Image Processing," Proceedings of DARPA Workshop, Nov. 1978, Pittsburgh, PA, pp. 191-195.

MISSION
of
Rome Air Development Center

RADC plans and executes research, development, test and selected acquisition programs in support of Command, Control Communications and Intelligence (C^3I) activities. Technical and engineering support within areas of technical competence is provided to ESD Program Offices (POs) and other ESD elements. The principal technical mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, ionospheric propagation, solid state sciences, microwave physics and electronic reliability, maintainability and compatibility.